

TARTU ÜLIKOOL  
LOODUS- JA TEHNOLOOGIA TEADUSKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT  
EVOLUTSIOONILISE BIOLOOGIA ÕPPETOOL

Helen Post

**Y-kromosomaalse haplogrupi N osa uurali keeli kõnelevate  
populatsioonide geenitiigi kujunemisel**

Magistritöö

Juhendajad:

Ph.D. Maere Reidla

Ph.D. Siiri Rootsi

Tartu 2014

# SISUKORD

<b>KASUTATUD LÜHENDID .....</b>	<b>4</b>
<b>SISSEJUHATUS.....</b>	<b>5</b>
<b>1. KIRJANDUSE ÜLEVAADE .....</b>	<b>6</b>
1.1. Inimese Y-kromosoomi iseloomustus .....	6
1.2. Fülogeneetilistes uuringutes kasutatavad Y-kromosoomi markerid ja mutatsioonikiiruse hindamine.....	7
1.3. Y-kromosoomi fülogeneesipuu.....	10
1.4. Uurali keeli kõnelevate populatsioonide iseloomustus.....	11
1.4.1. Populatsioonide iseloomustus keelte ja geograafilise paiknemise alusel.....	11
1.4.2. Populatsioonide iseloomustus Y-kromosomaalsete haplogruppide sageduste põhjal .....	15
1.5. Y-kromosoomi haplogruppide N1b ja N1c lühiiseloomustus .....	18
1.6. Y-kromosoomi mitterekombineeruva regiooni järjestuste kasutamine fülogeneetilistes ja –geograafilistes uuringutes .....	20
<b>2. EKSPERIMENTAALOSA.....</b>	<b>22</b>
2.1. Töö eesmärgid.....	22
2.2. Materjal ja metoodika .....	22
2.2.1. Töös kasutatud proovid .....	22
2.2.2. DNA amplifikatsioon .....	24
2.2.3. RFLP analüüs .....	25
2.2.4. Sekveneerimine .....	25
2.2.5. STR analüüs.....	27
2.3. Praimerite disainimine .....	28
2.4. Andmeanalüüs .....	29
<b>3. TULEMUSED JA ARUTELU .....</b>	<b>31</b>
3.1. Y-kromosoomi haplogrupi N fülogeneesipuu .....	31
3.2. Y-kromosoomi haplogrupi N1b fülogeograafia .....	34

3.3. Y-kromosoomi haplogrupi N1b haplotüüpide fülogeneetiline võrgustik.....	37
3.4. Nganassaanide ja neenetsite Y-kromosomi haplogrupiline koosseis .....	39
<b>KOKKUVÕTE .....</b>	<b>43</b>
<b>SUMMARY .....</b>	<b>45</b>
<b>TÄNUAVALDUSED .....</b>	<b>47</b>
<b>KASUTATUD KIRJANDUS.....</b>	<b>48</b>
<b>KASUTATUD VEEBIAADRESSID .....</b>	<b>56</b>
<b>LISAD .....</b>	<b>57</b>
LISA 1.....	57
LISA 2.....	60
LISA 3.....	62
Lisades kasutatud artiklid .....	67

## KASUTATUD LÜHENDID

ap	aluspaar
BSP	<i>Bayesian Skyline Plot</i>
GTR	<i>general time-reversible</i>
MSY	Y-kromosoomi meessoos-spetsiifiline regioon ( <i>male-specific region of the Y chromosome</i> )
mtDNA	mitokondriaalne DNA
NRX	Y-kromosoomi mitterekombineeruv regioon ( <i>non-recombining region of the Y chromosome</i> )
PAR	pseudoautosomaalsed regioonid
PCA	põhikomponentanalüüs ( <i>Principal Component Analysis</i> )
PCR	polümeraasne ahelreaktsioon ( <i>Polymerase Chain Reaction</i> )
RFLP	restriksioonifragmentide pikkuse polümorfism ( <i>Restriction Fragment Length Polymorphism</i> )
SNP	ühenukleotiidne polümorfism ( <i>single nucleotide polymorphism</i> )
STR	lühikesed tandeemsed kordusjärjestused ( <i>short tandem repeats</i> )
VUR	Volga-Uurali regioon
YCC	<i>Y Chromosome Consortium</i>
ISOGG	<i>International Society of Genetic Genealogy</i>

## SISSEJUHATUS

Mitukümmend aastat on fülogeneetilistes ja –geograafilistes uuringutes kasutusel olnud Y-kromosoom, mis pärandub vaid isaliinipidi (isalt pojale). Suure osa Y-kromosoomist moodustab X-kromosoomiga mitterekombineeruv meesso-spetsiifiline regioon. Y-kromosoomi fülogeneetilistes uuringutes on peamiselt kasutatud meesso-spetsiifilises piirkonnas esinevaid ühenukleotiidseid polümorfisme ja mikrosatelliite. Tänu uute ja paremate sekveneerimismeetodite väljatöötamisele ja rakendamisele on fülogeneetilistes uuringutes kasutusele võetud Y-kromosoomi mitterekombineeruva osa järjestused, mille tõttu on paranenud võimalused Y-kromosoomi fülogeneesipuu topoloogia ja Y-kromosoomi mutatsioonikiiruse täpsustamiseks ning uute bialleelsete markerite leidmiseks fülogeograafilistele analüüsidele.

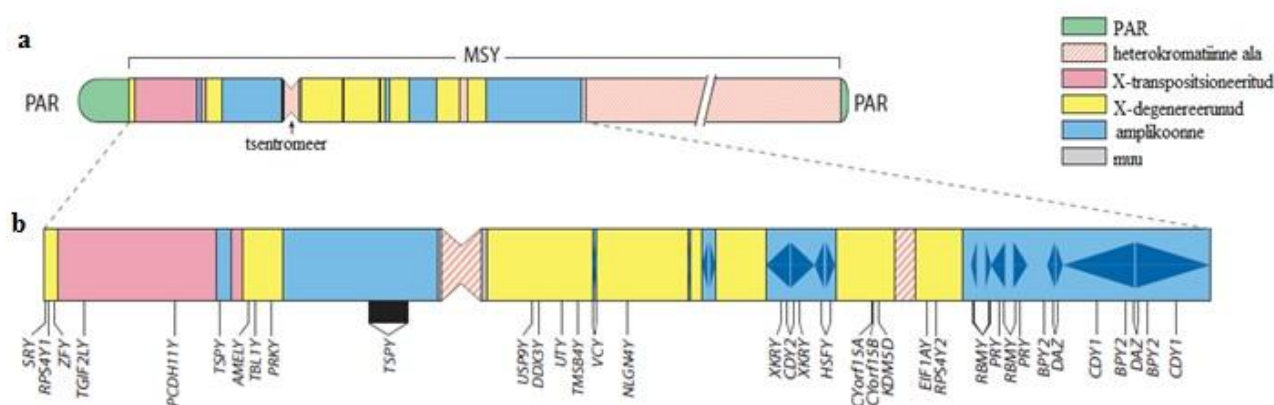
Kagu-Aasia mandriosas tekkinud haplogrupp N on tänapäeval levinud Euraasia põhjaosas, ulatudes Kaug-Idast Skandinaaviani. Laiaulatuslikku piirkonda haplogrupi N levialast asustavad uurali keeli kõnelevad rahvad. Varasemad madalal resolutsioonitasemel tehtud Y-kromosoomi analüüsid näitasid, et Siberi ning Ida-Euroopa populatsioonides on enam esindatud haplogrupi N alamhaplogrupid N1b ning N1c.

Käesoleva töö eesmärgiks on täpsustada haplogrupi N fülogeneesipuud Y-kromosoomi mitterekombineeruva osa järjestuste alusel ja põhjalikumalt uurida alamhaplogrupi N1b fülogeograafiat, kasutades uusi binaarseid markereid, mis on välja valitud Y-kromosoomi järjestustest. Töös kirjeldatakse haplogrupi N osa uurali keeli kõnelevate rahvaste geenitiigi kujunemisel ning täpsemalt iseloomustatakse kahe uurali keeli kõneleva populatsiooni – nganassaanide ja neenetsite – Y-kromosoomi haplogrupilist koosseisu.

# 1. KIRJANDUSE ÜLEVAADE

## 1.1. Inimese Y-kromosoomi iseloomustus

Inimese Y-kromosoom (joonis 1) on haploidne ja koosneb meessoo-spetsiifilisest regioonist (MSY, *male-specific region of the Y chromosome*) e mitterekombineeruvast alast (NRY, *non-recombining region of the Y chromosome*) (Jobling and Tyler-Smith, 2003), millega külgnevad kromosoomi otstes asuvad X-kromosoomiga homoloogsed pseudoautosomaalsed regioonid (PAR). NRY-s paiknevad heterokromatiinsed alad on kokku umbes 40 megaaluspaari pikad (Bachtrog, 2013). NRY eukromatiinse ala moodustavad amplikoonsed segmendid, X-transpositsioneeritud ja X-degenereerunud piirkonnad. Amplikoonsed segmendid sisaldavad palju kordusjärjestusi ning palindroomseid piirkondi. X-transpositsioneeritud ala on umbes 3,4 megaaluspaari pikk ja esineb ainult inimese Y-kromosoomis (Skaletsky *et al.*, 2003, Bachtrog, 2013).



**Joonis 1.** Inimese Y-kromosoomi skemaatiline kujutis. Joonise a osas on näidatud kogu inimese Y-kromosoomi skemaatiline kujutis. Joonise b osas on näidatud meessoo-spetsiifilise ala eukromatiinne piirkond. Y-kromosoomi eri piirkonnad on märgitud erinevate värvidega: rohelisega on märgitud pseudoautosomaalsed regioonid (PAR), heledama roosaga heterokromatiinne ala, tumedama roosaga X-transpositsioneeritud piirkonnad, kollasega X-degenereerunud ala ja sinisega amplikoonsed segmendid. Sinised kolmnurgad näitavad palindroomseid piirkondi ja vertikaalsete joontega on märgitud valke kodeerivate geenide asukohad (Hughes & Rozen, 2012).

Y-kromosoomi meessoo-spetsiifilist regiooni iseloomustab meiootilise rekombinatsiooni puudumine, mille tõttu pärandub mitterekombineeruv osa muutumatuna isalt pojale. NRY varieerub vaid tänu aja jooksul tekkivatele ning kuhjuvatele mutatsioonidele (Jobling & Tyler-Smith, 2003).

Y-kromosoomi haploidsus mõjutab Y-kromosoomi efektiivse populatsiooni suurust. Eeldades, et sugude osakaal on 1:1, on Y-kromosoomi efektiivse populatsiooni suurus 1/4 autosoomide efektiivse populatsiooni suurusest (Jobling & Tyler-Smith, 2003). Väikese efektiivse populatsiooni tõttu on geenitriivi toime Y-kromosoomile palju tugevam kui diploidsetele kromosoomidele, mille tõttu on Y-kromosoomi haplogruppide levikumuster enamasti selgemalt välja joonistunud kui autosomaalsete markerite korral (Jobling & Tyler-Smith, 2003).

Y-kromosoomi haplogruppide geograafilist levikut on mõjutanud ka asjaolu, et umbes 70% tänapäeva populatsioonidest on patrilokaalsed, s.o rohkem mehi, võrreldes naistega, jääb elama oma sünnikoha lähedale (Burton *et al.*, 1996, Jobling & Tyler-Smith, 2003). Seetõttu on isaliinide levik lokaalselt diferentseerunud kui emaliinidel (Jobling & Tyler-Smith, 2003). Matrilokaalsetes populatsioonides on vastupidi – Y-kromosoomi haplogruppide jaotus on vähem liigestunud kui mtDNA haplogruppide levik (Oota *et al.*, 2001).

## **1.2. Fülogeneetilistes uuringutes kasutatavad Y-kromosoomi markerid ja mutatsioonikiiruse hindamine**

Y-kromosoomi fülogeneetilistes uuringutes on peamiselt kasutatud bialleelseid markereid nagu ühenukleotiidsed polümorfismid (*single nucleotide polymorphism* - SNP), deletsioonid ja insertsioonid ning multialleelseid markereid nagu mikrosatelliidid e lühikesed kordusjärjestused (*short tandem repeats* - STR) tavaliselt pikkusega 2-6 aluspaari (de Knijff, 2000, Underhill & Kivisild, 2009). Bialleelsete markerite alleelivariantide kombinatsioonide põhjal määratletakse haplogrupid. Haplogruppide sisemise mitmekesisuse iseloomustamiseks kasutatakse STR-e. Haplogrupi sees olevaid alamliine, mida defineerivad STR-de alleelide kombinatsioonid, nimetatakse haplotüüpideks (de Knijff, 2000). SNP-dele omane madal mutatsioonikiirus võimaldab uurida inimpopulatsioonide varasemaid demograafilisi protsesse (YCC, 2002) ja STR-de suurem kiirus suhteliselt hiljutisi demograafilisi sündmusi (de Knijff, 2000).

Varasemates töödes on haplogruppide ekspansioonaja hindamisel sageli kasutatud STR-e. STR-dele on meetodist sõltuvalt saadud erinevaid mutatsioonikiiruse hinnanguid. Sugupuude uuringutel leitud mutatsioonikiirus:  $2,0 \cdot 10^{-3}$  lookuse kohta generatsiooni kohta (Heyer *et al.*, 1997) erineb suuresti efektiivsest mutatsioonikiirusest  $6,9 \cdot 10^{-4}$  lookuse kohta generatsiooni kohta, mis on arvatud kolme andmekogu põhjal, võttes arvesse ainult fikseerunud mutatsioone (Zhivotovsky *et al.*, 2004). Erinevused tekivadki sellest, et sugupuude põhjal arvatud mutatsioonikiirused võtavad arvesse kõik mutatsioonid, aga STR-

de efektiivne mutatsioonikiirus arvestab vaid populatsioonis fikseerunud mutatsioone. Evolutsioonilistes uuringutes ja populatsioonigeneetikas on õigustatud efektiivse mutatsioonikiiruse kasutamine, sest populatsioonis kõik mutatsioonid ei fikseeru (STR-de variatsioon väheneb) (Zhivotovsky *et al.*, 2006).

SNP-del on oluliselt väiksem mutatsioonikiirus kui STR-del. Y-kromosomaalse DNA mutatsioonimäär  $0,82 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Poznik *et al.*, 2013) on suurem kui tuumagenoomil, kuid oluliselt väiksem kui mitokondriaalsel DNA-l (Graves, 2006, Cruciani *et al.*, 2011). Viimastel aastatel on uuesti tõstatatud Y-kromosoomi mutatsioonikiiruse küsimused, kuna tänu uutele sekveneerimistehnoloogiatele on üha suuremal hulgal indiviididel sekveneeritud järjest pikemad piirkonnad Y kromosoomist. Siiski ei suudeta praeguste tehnoloogiatega sekveneerida heterokromatiinseid piirkondi ning järjestada pikki kordusjärjestusi ja palindroomse sisaldavaid amplikoonseid regioone ning suure homoloogia tõttu X-kromosoomiga on keeruline ka X-transpositsioneeritud piirkondade analüüs (Poznik *et al.*, 2013). Kuna Y-kromosoomist moodustavad suure osa heterokromatiinsed alad ning kordused ja palindroomid, saab mutatsioonikiiruste leidmiseks kasutada suhteliselt väikest osa kogu kromosoomist, täpsemalt X-degenereerinud piirkonda (Poznik *et al.*, 2013, Francalacci *et al.*, 2013, Mendez *et al.*, 2013). Inimese Y-kromosoomi jaoks on viimasel ajal avaldatud mitu mõnevõrra erinevat mutatsioonikiiruse hinnangut, mis põhinevad Y-kromosoomi järjestuste analüüsil:

1.  $1 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Xue *et al.*, 2009, Wei *et al.*, 2012) ehk üks asendus 100 aasta kohta,
2.  $0,82 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Poznik *et al.*, 2013) ehk 1 asendus 122 aasta kohta,
3.  $0,65 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Francalacci *et al.*, 2013) ehk 1 asendus 154 aasta kohta,
4.  $0,60 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Mendez *et al.*, 2013) ehk 1 asendus 162 aasta kohta.

Esimene eespool toodud mutatsioonikiiruse hinnangutest (Xue *et al.* 2009) põhineb ühe Hiina perekonna kahe teineteisest 13 generatsiooni kaugusel oleva liikme Y-kromosoomi järjestustel. Kasutati 10,15 megaaluspaari pikkust lõiku meessoos-spetsiifilisest regioonist. Arvestamata jäeti palindroomsed ja kordusjärjestused. Mutatsioonide arvu leidmiseks võrreldi



järjestusi Y-kromosoomi referentsjärjestusega. Antud mutatsioonikiiruse hinnang on suhteliselt ebatäpne (95%-lised usalduspiirid  $3 \cdot 10^{-10}$  -  $2.5 \cdot 10^{-9}$ ), mis on tingitud uuringus leitud väikesest mutatsioonide arvust. Xue poolt leitud mutatsioonikiirust kasutas Wei (Wei *et al.*, 2012) Y-kromosoomi fülogeneesipuu harudele ajahinnagute saamiseks. Selleks sekveneeriti maailma eripaigust pärineva 36 indiviidi Y-kromosoomid. Uuringus kasutati meessoospetsiifilise piirkonna järjestusi pikkusega 8,79 meegaaluspaari. Analüüsi kitsaskohaks on väike indiviidide hulk.

Poznik (Poznik *et al.* 2013) kasutas uuringus 69 erinevatest populatsioonidest pärit indiviidi Y-kromosoomi järjestusi, mis hõlmasid 9,9 miljonit aluspaari X-degeneratiivsest piirkonnast. Kalibreerimispunktiks võeti Ameerika asustamine (umbes 15 tuhat aastat tagasi) ning mutatsioonikiirust hinnati Q-M3 haplogruppi kuuluvate järjestuste põhjal. Kuna haplogruppi Q-M3 leidub nii Põhja- kui ka Lõuna-Ameerikas, eeldati, et see tekkis enne migratsiooni Lõuna-Ameerikasse.

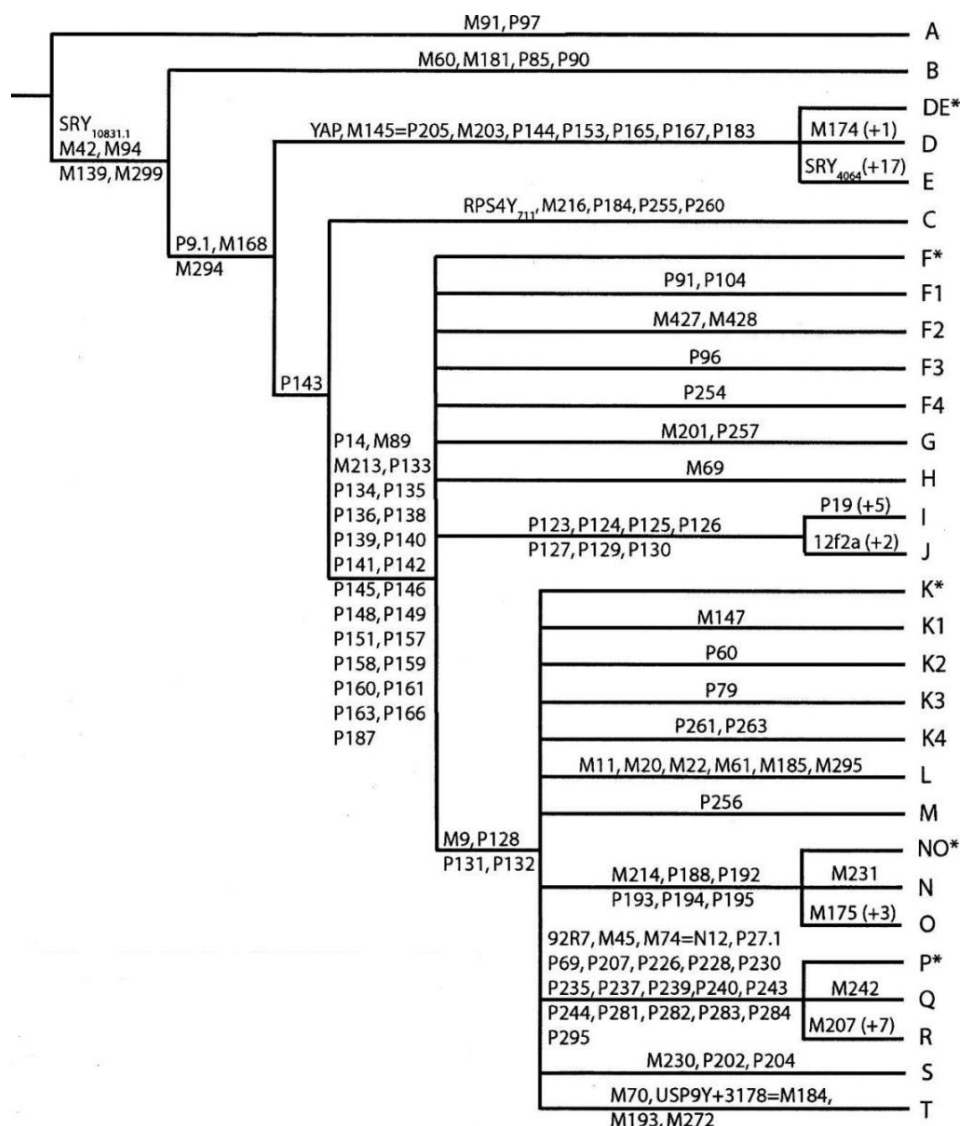
Francalacci (Francalacci *et al.*, 2013) töögrupp kasutas mutatsioonikiiruse hindamisel 1204 sardiinlase Y-kromosoomi järjestusi, mis hõlmasid meessoospetsiifilisest regioonist 8,79 meegaaluspaari. Kalibreerimiseks kasutati populatsiooni esmast ekspansiooni umbes 7700 aastat tagasi ning sardiinlastele iseloomuliku haplogrupi I2a1a 435 järjestust. Võrreldes ülejäänud uuringutega on antud töös kasutatud madala sekveneerimiskatvusega (4,32 kordne) järjestusi. Lisaks võib selles töös kasutatud kalibreerimisaeg (7700 aastat tagasi) olla ebatäpne, kuna põhineb arheoloogilistel leidudel, mis võivad olla mittetäielikud ning hõlmavad tihti suurt ajaintervalli.

Mendez (Mendez *et al.* 2013) tuletas Y-kromosoomi mutatsioonikiiruse autosoomide mutatsioonikiirusest. Kasutades sel moel leitud mutatsioonikiirust inimese Y-kromosoomi viimase ühise eellase vanuse arvutamisel, saadi tulemuseks 388000 aastat, mis on oluliselt vanem varasemast vanusehinnangust (141500 aastat (Cruciani *et al.*, 2011)) ja ka mtDNA viimase ühise eellase vanuse hinnangust (99000 aastat (Poznik *et al.*, 2013)). Lisaks ületas see 140000 aastaga anatoomiliselt tänapäevase inimese vanima leitud kolju vanuse (196000 aastat (McDougall *et al.*, 2005)). Need faktid näitavad, et Mendez on Y-kromosoomi tegelikku mutatsioonikiirust alla hinnanud. Lisaks, kasutati selles uuringus vanuse hindamisel erineva pikkusega Y-kromosoomi järjestusi (haplogruppi A00 kuuluv järjestus 240 tuhat ja haplogruppi A0 esindav järjestus 180 tuhat aluspaari pikk), mis ka vanusehinnangut mõjutas (Elhaik *et al.*, 2014).

Eelmainitud mutatsioonikiiruse hinnangutest on kõige usalduseväärsem  $0,82 \cdot 10^{-9}$  asendust nukleotiidi kohta aastas (Poznik *et al.*, 2013), kuna antud uuringus kasutati võrdlemisi suure sekveneerimiskatvusega järjestusi ning analüüsi kaasatud mutatsioonide arv oli samuti suur.

### 1.3. Y-kromosoomi fülogeneesipuu

Y-kromosoomi haplogrupid on binaarsete markerite omandatud (*derived*) alleelide alusel defineeritud NRY liinid (YCC, 2002). 2002. aastal avaldas *Y Chromosome Consortium* (YCC) Y-kromosoomi fülogeneesipuu 153 haplogrupiga, mida defineeriti 243 binaarse markeriga (YCC, 2002). 2003. aastal avaldas YCC veelgi täpsema fülogeneesipuu (Jobling & Tyler-Smith, 2003). Hiljem täiendasid puud Karafet (Karafet *et al.*, 2008) (joonis 2) ja van Oven (van Oven *et al.*, 2013). Y-kromosoomi fülogeneesipuu topoloogia täiendamine on hetkel väga aktuaalne ja seetõttu täpsustub puu pidevalt – kõige kiiremini täiendatakse ISOGG (*International Society of Genetic Genealogy*) Y-kromosoomi fülogeneesipuud (<http://www.isogg.org/tree>). Selline kiire täiendamine on juurde toonud palju uusi markereid, mis võimaldavad Y-kromosoomi täpsemalt genotüpiseerida. Samas on selline olukord muutnud Y-kromosoomi haplogruppide nimed pikkadeks ja kohmakateks. Kuna paljudes fülogeograafilistes uuringutes kasutatakse Karafeti (Karafet *et al.*, 2008) esitatud puu topoloogiat, siis on sellest lähtunud ka käesolevas töös.



**Joonis 2.** Y-kromosoomi markerite fülogeneesipuu skemaatiline kujutis. Harudele on märgitud defineerivad mutatsioonid ning harude ottesse on märgitud haplogruppide nimetused (Karafet *et al.*, 2008)

## 1.4. Uurali keeli kõnelevate populatsioonide iseloomustus

### 1.4.1. Populatsioonide iseloomustus keelte ja geograafilise paiknemise alusel

Uurali keelte lai leviala ulatub Lääne-Siberist Ida-Euroopasse (joonis 3). Kõige läänepoolsem uurali keeli kõnelev populatsioon on ungarlased, kes elavad Kesk-Euroopas. Uurali keelkond jaguneb kaheks: samojeedi keeled ja soome-ugri keeled (joonis 4 ja joonis 5) (Honkola *et al.*, 2013; Syrjänen *et al.*, 2013). Samojeedi keeli räägivad nganassaanid, eenetsid, neenetsid ja sõlkupid. Soome-ugri keelte sügavama haru moodustavad erinevate keelteaduslike koolkondade teooriate kohaselt kas ugri keeled (Honkola *et al.*, 2013) või permi keeled (Syrjänen *et al.*, 2013). Ugri keeli räägivad ungarlased, mandsid ja handid ning

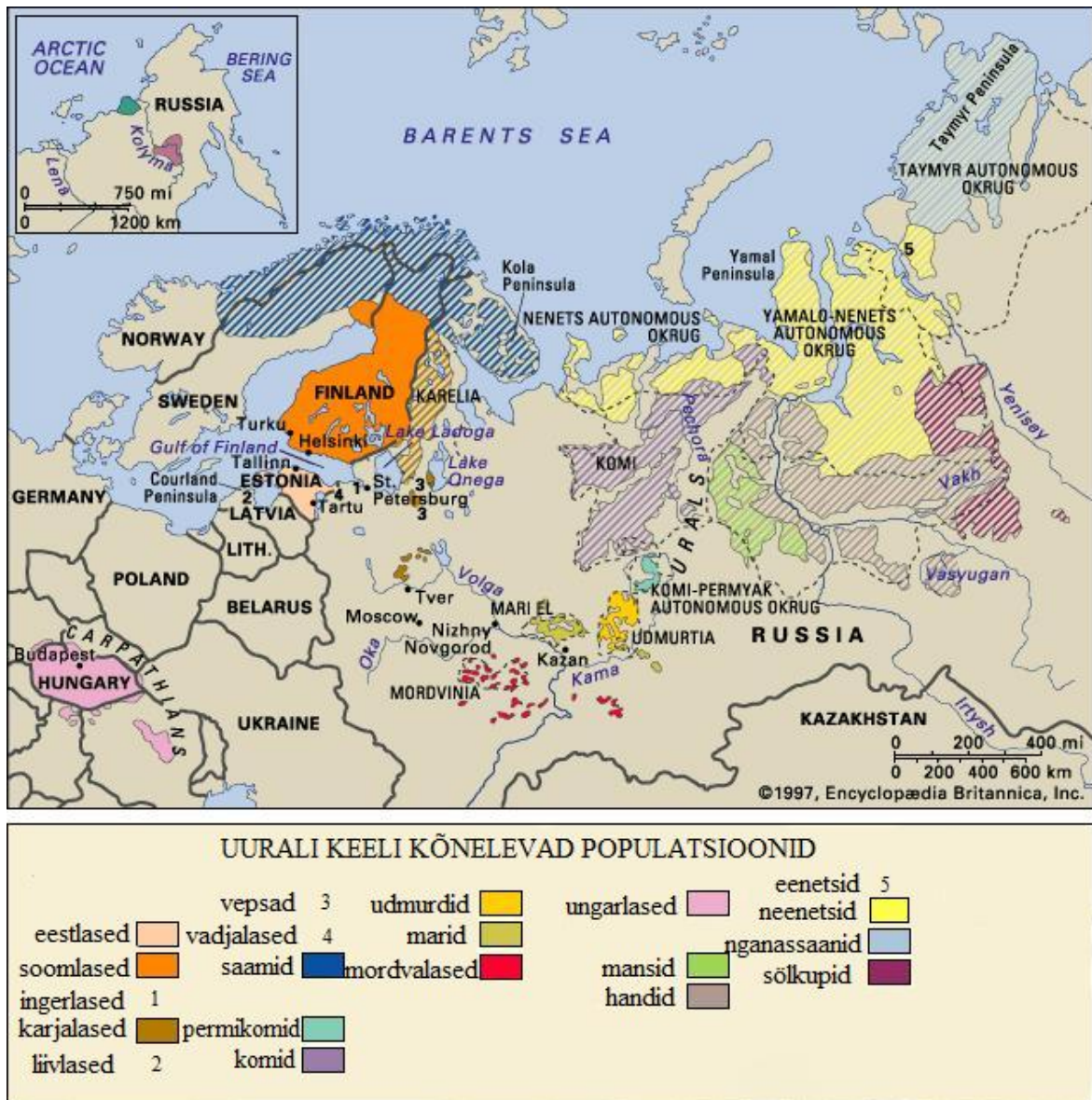
permi keelte hulka kuuluvad udmurdi ja komi keeled. Noorimasse soome-volga harusse kuuluvad mordva, mari, saami, soome, eesti, karjala ja vepsa keeled.

Nganassaanid elavad Loode-Siberis Taimõri poolsaarel ja kõnelevad samojeedi keelte hulka kuuluvat nganassaani keelt. Tegu on Euraasia kõige põhjapoolsema põlisrahvaga, kelle põhiliseks tegevusalaks on põhjapõdrakasvatus aga ka kalapüük. Nganassaanid jagunevad kaheks etniliseks grupiks: avamid läänes ja vadejevid idas (Golovnev, 1999). Etnilise populatsiooni suurus Venemaal on umbes 800 (<http://www.ethnologue.com>), kellest vaid väike osa kõneleb nganassaani keelt ladusalt ja seetõttu loetakse rahvust väljasuremisohus olevaks (<http://www.eki.ee/books/redbook/>). Nganassaanidest läänes elavad samuti samojeedi keeli rääkivad eenetsid ja neenetsid. Idas ja lõunas piirnevad nganassaanid evenkide ja dolgaanidega (Golovnev, 1999), kes kõnelevad altai keelkonda kuuluvaid keeli (<http://www.ethnologue.com>).

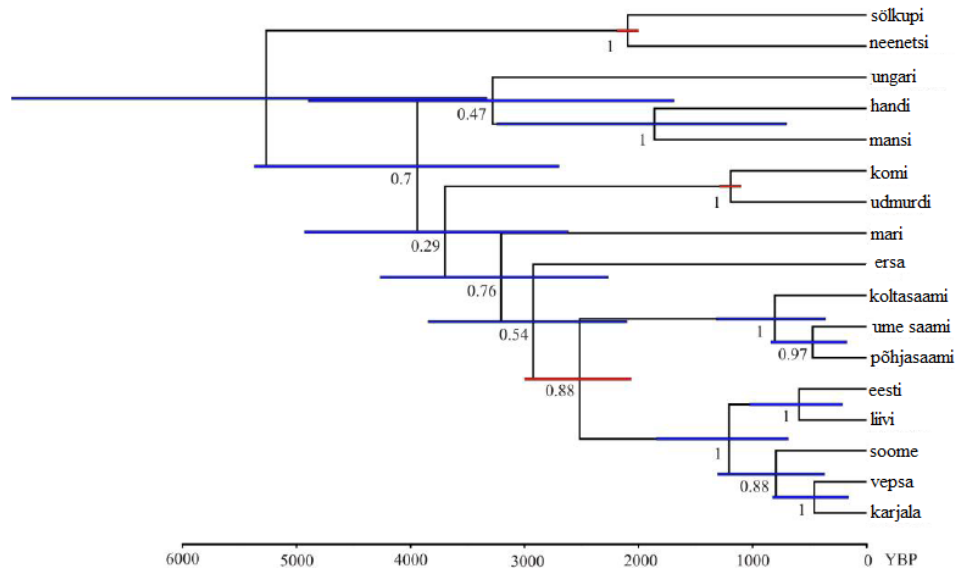
Sarnaselt nganassaanidele olid neist läänes elavad neenetsid algselt nomaadirahvas, kes tegelesid põhjapõdrakasvatusega. Neenetsid jagunevad elupaikade järgi metsa ja tundra neenetsiteks (Golovnev & Osherenko, 1999). Populatsiooni suurus Venemaal on umbes 41 000 (<http://www.ethnologue.com>). Lähimateks naabriteks on nganassaanid ja eenetsid Taimõri poolsaarel. Lõunas on naabriteks handid, mansid ja sõlkupid, kes elavad Obi jõe äärses taigas, ning läänes sürjakomid, kes asustavad Kirde-Euroopa tundraalasid. Tänapäeval elavad selles piirkonnas ka venelased ja ukrainlased, kes kõnelevad indoeuroopa keelkonda kuuluvaid slaavi keeli (Golovnev & Osherenko, 1999).

Volga jõe ja Uurali mäestiku vahelisel alal nn Volga-Uurali regioonis (VUR) elavad mitmed uurali keeli kõnelevad populatsioonid: permikomid, marid, udmurdid ning mordvalased. Komisid on 2010. aasta rahvaloenduse järgi umbes 220 000 (<http://www.ethnologue.com>). Geograafilise asendi ning väikeste keeleliste erinevuste tõttu jagunevad komid permikomideks ning sürjakomideks. Algselt elatusid komid jahist ning kalapüügist, hiljem tulid tegevusaladena juurde ka põlluharimine ning põhjapõdrakasvatus (Olson, 1994). Keeleliselt on komidele kõige lähedasemad udmurdid, kelle põhilisteks tegevusaladeks on olnud põlluharimine ning karjakasvatus. Lisaks käidi jahil ning tegeleti mesindusega, mis on oluline tegevusala mitmetel Volga-Uurali piirkonna rahvastel (Cole, 2011). Volga-Uurali regioonis elavad ka venelased ning altai keeli kõnelevad tatarlased, baškiirid ja tšuvašid.

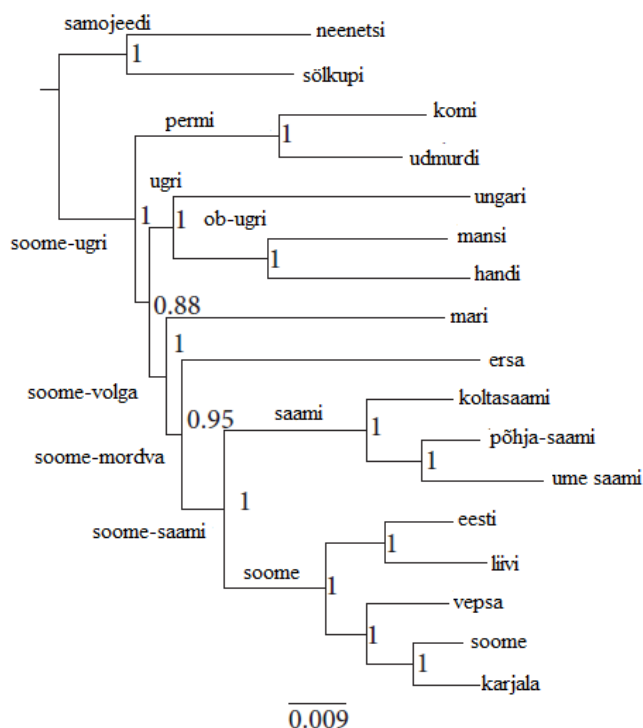
Skandinaavia poolsaarel elavad uurali keeli kõnelevad saamid on võrdlemisi väike populatsioon umbes 30000 indiviidi (<http://www.ethnologue.com>). Läänemere ääres elavad rahvaarvult samuti väikesed vepsad umbes 4000-5000 indiviidi ja karjalased umbes 30000 indiviidi (<http://www.ethnologue.com>). Samas piirkonnas elavad ka eestlased ja soomlased, kelle populatsioonid on suuremad vastavalt 1040000 ja 5100000 (<http://www.ethnologue.com>).



**Joonis 3.** Uurali keeli kõnelevad rahvad. Viirutatud alad tähistavad hõredalt asustatud alasid (<http://www.britannica.com>).



**Joonis 4.** Uurali keelte fülogeneesipuu. Puu on konstrueeritud 226 põlissõnavarasse kuuluva sõna ja nende tähenduste põhjal. Sõlmede juures on näidatud rühmade aposterioorsed tõenäosused. Väärtused 0,95-1 näitavad tugevat ning väärtused, mis jäävad alla 0,9, madalat statistilist toetust. Sinised ning punased horisontaaljooned näitavad divergentsiaegade 95%-lisi usaldusvahemikke. (Honkola *et al.*, 2013).



**Joonis 5.** Uurali keelte fülogeneesipuu. Puu on konstrueeritud 226 põlissõnavarasse kuuluva sõna ja nende tähenduste põhjal. Sõlmede juures on näidatud rühmade aposterioorsed tõenäosused. Väärtused 0,95-1 näitavad tugevat ning väärtused, mis jäävad alla 0,9, madalat statistilist toetust. Harude pikkused näitavad muutuste hulka (Syrjänen *et al.*, 2013).

### 1.4.2. Populatsioonide iseloomustus Y-kromosomaalsete haplogruppide sageduste põhjal

Iseloomustamiseks Lääne-Siberis ja Ida-Euroopas elavaid uurali keeli kõnelevaid populatsioone ja nende naabreid Y-kromosomaalsete haplogruppide sageduste põhjal, tehti kirjanduses avaldatud andmete (Semino *et al.*, 2000, Wells *et al.*, 2001, Karafet *et al.*, 2002, Passarino *et al.*, 2002, Sanchez *et al.*, 2003, Rootsi *et al.*, 2004, Tambets *et al.*, 2004, Kharkov *et al.*, 2005, Derenko *et al.*, 2006, Karlsson *et al.*, 2006, Lappalainen *et al.*, 2006, Balanovsky *et al.*, 2008, Lappalainen *et al.*, 2008, Pimenoff *et al.*, 2008, Dulik *et al.*, 2011, Dulik *et al.*, 2012, Fedorova *et al.*, 2013) alusel põhikomponentanalüüs (PCA), kasutades selleks programmi POPSTR (<http://harpending.humanevo.utah.edu/popstr>) (joonis 6).

Jooniselt on näha, et enamasti klasterduvad uurali keeli kõnelevad populatsioonid kokku oma geograafiliste naabritega, jagades nendega ühiseid Y-kromosomaalseid haplogruppe. Ida-Euroopas k.a Volga-Uurali regioonis (VUR) elavatel soome-ugrilastel on esindatud suhteliselt palju erinevaid Y-kromosoomi haplogruppe. Üheks enam esindatud haplogrupiks on haplogrupp R1, mille alamharusid R1a ja R1b leidub palju Euroopa populatsioonides, näiteks rootslastel, norralastel, poolakatel ning kesk- ja lõunavenelastel. R1a aja R1b esinevad Euroopas suure sagedusega ja vastupidise, ida-lääne suunalise gradiendiga. R1b on valdav Lääne- ja Põhja-Euroopas ning sagedus väheneb ida suunas. R1a moodustab kuni poole paljude Ida-Euroopa rahvaste (eriti slaavi keeli kõnelevate rahvaste) geenitiigist ja sagedus kahaneb lääne suunas. Haplogrupp R1a on levinud ka Lõuna-Siberis elavatel altai keelkonda kuuluvatel populatsioonidel (kuni 40%) (Underhill *et al.*, 2014). R1a subhaplogrupid R1a-458 ja R1a-M558 on sarnase levikuga (suurima sagedusega Ida- ja Kesk-Euroopa populatsioonidel), kuid R1a-558 on levinud ka Volga-Uurali regioonis. Kesk- ja Lõuna-Siberi populatsioonides neid klaade peaaegu ei esine. Samas on sealsetel rahvastel nagu näiteks altailastel ja hakassidel sage R1a-Z93 (üle 30%) (Underhill *et al.*, 2014).

Teiseks väga sagedaseks haplogrupiks on haplogrupp N, mille sagedus on kõrge ka paljudes Siberi populatsioonides. Ida-Siberis elavatel jakuutidel ulatub N1c osakaal isegi 95 protsendini (Fedorova *et al.*, 2013). N1c suure sageduse tõttu klasterduvad jakuudid PC analüüsil kokku Volga-Uurali piirkonna populatsioonidega. Kirjanduses avaldatud andmete puhul on haplogrupeerimise lahutustase madal ning seetõttu on PCA lahutusvõime väike ja osade populatsioonide puhul võivad erinevused märkamata jääda. Lääne-Siberis elavates uurali keeli kõnelevates populatsioonides (mansid, neenetsid, nganassaanid) on sagedaseim haplogrupp N1b. Samuti leidub N1b haplogruppi VUR-i populatsioonides (komidel,

udmurditel ning veidi ka mordvalastel). Lõuna-Siberi populatsioonidest on N1b kõige sagedasem tõvadel ja hakassidel. Evenidel ja evenkidel on samuti N1b võrdlemisi sage, kuigi haplogruppi N1c leidub rohkem (Duggan *et al.*, 2013).

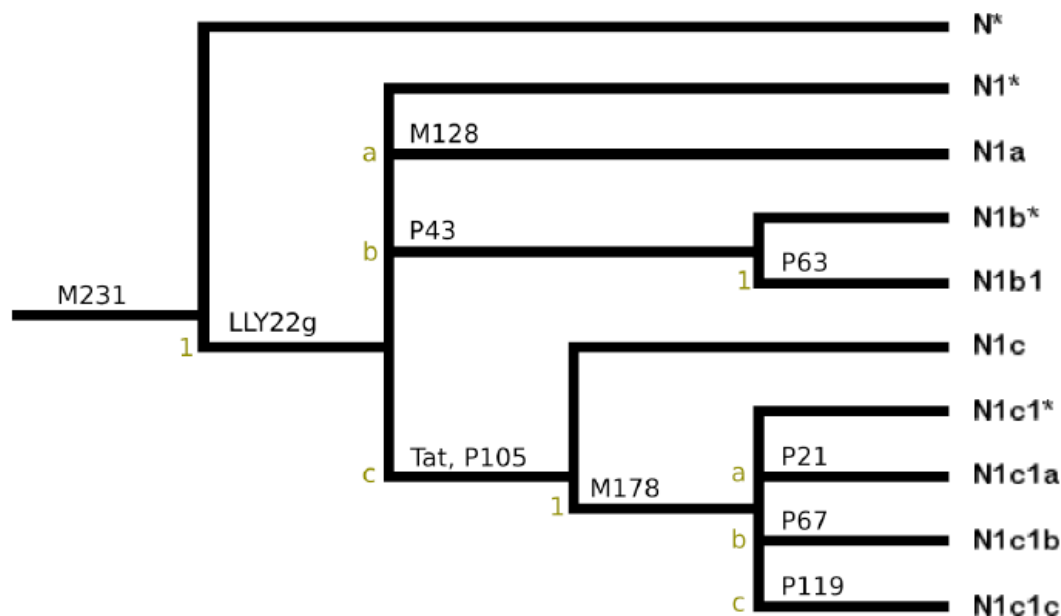
Lääne-Siberis elavad sõlkupid eristuvad ülejäänud uurali keeli kõnelevatest populatsioonidest haplogrupi Q suure sageduse tõttu (~66%). Haplogrupp Q on valdav sõlkupite naabritel ketidel (~93%) ning sage ka Lõuna-Siberi populatsioonidel (altailastel ja tõvadel).



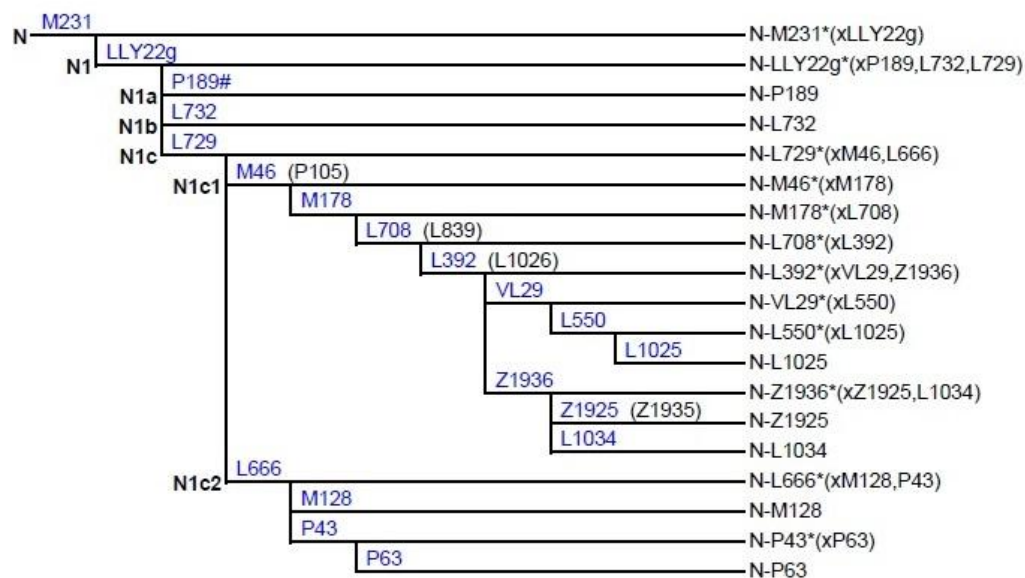


## 1.5. Y-kromosoomi haplogrupid N1b ja N1c lühiiseloostus

Y-kromosoomi fülogeneesipuu täiendamiste käigus on mitmekesistunud ka N haplogrupp – juurde on tulnud mitmeid alamklaade, sealhulgas ka N1c ja N1b (Karafet *et al.*, 2008, van Oven *et al.* 2013) (joonis 7 ja joonis 8).



**Joonis 7.** N haplogrupi fülogeneesipuu. Harudele on märgitud defineerivad mutatsioonid ning harude tippudes on haplogrupi nimetused (väljalõige, Karafet *et al.*, 2008).



**Joonis 8.** N haplogrupi fülogeneesipuu. Harudele on märgitud defineerivad mutatsioonid ning harude tippudes on haplogrupi nimetused (väljalõige, van Oven *et al.*, 2013)

Y-kromosoomi haplogrupi N alamhaplogruppidest on sagedasemad N1b (N2 YCC, 2002 järgi), mida defineerib marker P43 (YCC, 2002), ja N1c (N3 YCC, 2002 järgi), mis on määratletud markeriga TAT (Zerjal *et al.*, 1997, YCC 2002; Karafet *et al.*, 2008). Y-kromosomaalne haplogrupp N ja selle alamhaplogrupid N1b ja N1c on Siberi ja Ida-Euroopa populatsioonide hulgas kõige sagedasemad ning haruldased Kagu- ja Ida-Aasia mandriosas (Shi *et al.*, 2013).

Varasemad uuringud STR-ide ja SNP-d põhjal on näidanud, et N haplogrupp võis algselt levima hakata Kagu-Aasia mandriosast läänepoolse 12 kuni 14 tuhat aastat tagasi (Rootsi *et al.*, 2007). Seda arvamust toetab asjaolu, et Kagu-Aasia piirkonna populatsioonides on esindatud kõik haplogrupi N sügavamad harud (Rootsi *et al.*, 2007, Derenko *et al.*, 2007). Samas on arvatud ka, et N haplogrupp võis tekkida Lõuna-Hiinas 21 tuhat aastat tagasi, sealt laieneda Põhja-Hiinasse umbes 12-18 tuhat aastat tagasi ning siis levida põhja poole Siberi populatsioonidesse kuskil 12-14 tuhat aastat tagasi ning jõuda Kesk-Aasia ja Euroopa populatsioonidesse umbes 8-10 tuhat aastat tagasi (Shi *et al.*, 2013). Kagu-Hiinas Liao jõe orust neljast erinevast paigast kogutud vanaDNA analüüs näitas, et alamhaplogrupp N1-M231 oli selle piirkonna populatsioonides olemas enam kui 5000 aastat tagasi (Cui *et al.*, 2013). Antud uuringus määrati Y-kromosomaalsed haplogrupid 47 proovil ning leiti, et 5000 aastat tagasi oli sealsetes populatsioonides kõige sagedasem haplogrupp N1 (xN1a, N1c) (63%), kuid mitmete segunemiste tõttu lõunapoolt tulnud indiviididega on N1 sagedus selles piirkonnas praeguseks märkimisväärselt langenud (umbes 8%) (Cui *et al.*, 2013).

Kõige sagedasem alamklaad N1c tekkis ilmselt tänapäeva Hiinas umbes 12 tuhat aastat tagasi ning levis Siberisse ja edasi Ida-Euroopasse (Rootsi *et al.*, 2007). N1c on Siberis valdav, saavutades suurima sageduse jakuutidel (80-90%) (Rootsi *et al.*, 2007, Fedorova *et al.*, 2013). Haplogrupp N1c on laialt levinud ka Volga-Uurali regiooni populatsioonides udmurtidel, (Pimenoff *et al.*, 2008), komidel ning veidi vähem maridel (Rootsi *et al.*, 2007). N1c sagedus väheneb idast läände ja eriti märgatav on see Ida-Euroopa populatsioonides, kus sageduse langus lääne suunas on järsk (Rootsi *et al.*, 2007, Derenko *et al.* 2007). Mirabali (Mirabal *et al.*, 2009) väitel ei liikunud N1c idast läände vaid läänest itta, sest Ida-Euroopa populatsioonides (soomlased, eestlased, põhja-venelased) on N1c sagedus kõrgem kui neist läänes elavatel sürjakomidel. Seda teooriat aga ei toeta N haplogrupi teadaolev fülogenees ja ka asjaolu, et väikestes populatsioonides võivad haplogruppide sagedused drastiliselt muutuda ja väheneda STR-de mitmekesisus. Suurematel populatsioonidel esineb sagedamini ekspansiooniga seotud diversiteedi kasvu.

STR-ide variatsiooni järgi jagatakse N1c kahte rühma: üks, mis hõlmab Volga-Uurali regiooni populatsioone ja turgi keeli kõnelevaid populatsioone ning teine, mis on sagedasem eestlaste, karjalaste, vepsade, aga ka venelaste ning ukrainlaste seas (Rootsi *et al.*, 2007, Derenko *et al.* 2007).

Haplogrupi N teise alamklaadi N1b vanus võib olla kuni 16 tuhat aastat (Mirabal *et al.*, 2009). Antud haplogrupp on Siberis populatsiooniti väga erineva sagedusega, olles levinud uurali keeli kõnelevatel metsa ja tundra neenetsitel, nganassaanidel, eenetsitel, hantidel, mansidel, sõlkupitel, aga ka turgi keele kõnelejate hulka kuuluvatel dolgaanidel (Karafet *et al.*, 2002, Rootsi *et al.*, 2007, Pimenoff *et al.*, 2008, Duggan *et al.*, 2013). Suurima sageduse saavutab N1b nganassaanidel (92 %), eenetsitel (78%) ja tundra neenetsitel (74%) (Karafet *et al.*, 2002, Rootsi *et al.*, 2007). Võrreldes sõsarklaadiga N1c, on tema üldine levikupiirkond kitsam, piirdudes peamiselt Volga-Uurali regiooniga ja ei ulatu kaugemale vepsade asualast.

STR-ide mustri põhjal eristati N2-s (Rootsi *et al.*, 2007, Derenko *et al.*, 2007) (nüüd N1b (Karafet *et al.*, 2008)) Aasia alamklaad (N2-A) ja Euroopa alamklaad (N2-E). STR-de põhiste ajahinnangute alusel on näidatud, et Aasia alamklaad kujunes varem ja Euroopa haru tekkis Aasia omast (Rootsi *et al.*, 2007, Derenko *et al.*, 2007). Arvatakse, et Siberis toimus lääne suunas kaks migratsiooni: üks soome-ugri rahvaste migratsioon, mis jõudis Volga-Uurali piirkonda ja pani aluse N2-E tekkele, ning teine altai keelkonnna keeli rääkivate rahvaste migratsioon, mis pani aluse N2-A alamklaadi tekkele (Rootsi *et al.*, 2007).

## **1.6. Y-kromosoomi mitterekombineeruva regiooni järjestuste kasutamine fülogeneetilistes ja –geograafilistes uuringutes**

Sekveneerimismetoodika märkimisväärne areng on viinud inimese genoomi odavamale, kiiremale ning kvaliteetsemale järjestamisele. 2010. aastal kirjeldati uut *Complete Genomics* poolt välja töötatud sekveneerimisplatvormi (Drmanac *et al.*, 2010), mille abil sekveneeriti kolm inimese genoomi 45-87 kordse katvusega. Platvorm on võrdlemisi täpne – 1 viga 100 tuhande aluspaari kohta (Drmanac *et al.*, 2010). *Complete Genomics*-i platvormil sekveneeriti ka kolmest Aafrika küttide-korilaste populatsioonist pärit 15 indiviidi genoomid kogu ulatuses. 60-kordse katvusega genoomijärjestusi kasutati inimese evolutsioonilise mineviku rekonstrueerimisel (Lachance *et al.*, 2012). See oli üks esimesi populatsioonigeneetilisi uurimusi, mis kasutas analüüsis täisgenoomi järjestusi ning milles tõdeti, et endiste meetoditega ei oleks selliste tulemusteni jõutud (Lachance *et al.*, 2012).

Uus tehnoloogia on lihtsustanud ka Y-kromosoomi analüüsi ning viimasel ajal on aina enam sekveneeritud Y-kromosoomi meessoo-spetsiifilist regiooni (Cruciani *et al.*, 2011, Poznik *et al.*, 2013, Mendez *et al.*, 2013, Francalacci *et al.* 2013). Nimetatud uurimustöodes järjestati enam kui 1300 indiviidi Y-kromosoomi meessoo-spetsiifilisest regioonist umbes 8-10 miljonit aluspaari. 240 tuhande aluspaari pikkune lõik on sekveneeritud ka Y-kromosoomi fülogeneesipuu sügavamasse harusse (A00) kuuluvast Y kromosoomist (Mendez *et al.*, 2013).

Y-kromosoomi X-degenereerunud ala järjestused on muutunud kergemaks Y-kromosoomi fülogeneesipuu topoloogia täpsustamise ja haplogruppide vanuse hindamise. Uute suhteliselt pikkade järjestuste alusel arvutatud mutatsioonikiirused võimaldavad täpsemalt hinnata lahknemisi Y-kromosoomi fülogeneesipuul.

Meessoo-spetsiifilise regiooni sekveneerimine on samuti oluliselt lihtsustanud uute markerite leidmist. Varem uuriti avastatud mutatsiooni suurel hulgal proovidel ning tihti ei leitudki antud mutatsiooni teistel indiviididel, kuna tegu oli sellele ühele indiviidile omase mutatsiooniga. Sekveneerimise puhul on vaja vähemalt kahte järjestust (varasema suure hulga asemel), et välja selekteerida neile järjestustele ühised, n.n defineerivad mutatsioonid. Seejärel disainitakse praimerid ja kontrollitakse nende mutatsioonide olemasolu suuremal arvul indiviididel. Sel moel leitud markereid saab kasutada vastavate markeritega defineeritud kladide geograafilise leviku uurimiseks. Mõnedes fülogeneetilistes töodes on Y-kromosoomi järjestusi juba kasutatud; näiteks uuriti Askhenazi leviitide alamhaplogrupi paigutumist R1a fülogeneetilisele puule (Rootsi *et al.*, 2013). Selleks kasutati üheksat haplogruppi R1a kuuluvat proovi ning seitset R1b haplogruppi kuuluvat proovi (Rootsi *et al.*, 2013). Kaheksat haplogruppi R1a ja viite R1b järjestust kasutati divergentsiaja hindamiseks (Underhill *et al.*, 2014).

## **2. EKSPERIMENTAALOSA**

### **2.1. Töö eesmärgid**

Töö peamiseks eesmärgiks on täpsustada Y-kromosoomi haplogrupi N sisestruktuuri ja alamhaplogrupi N1b fülogeograafiat ning iseloomustada N1b osa uurali keeli kõnelevate populatsioonide geenitiigi kujunemisel.

- Y-kromosoomi X-degenereerunud ala järjestuste põhjal haplogrupi N fülogeneesipuu rekonstrueerimine.
- Alamhaplogrupi N1b kolme põhiklaadi geograafilise leviku selgitamine.
- Kahe uurali keeli kõneleva populatsiooni – nganassaanide ja neenetsite – Y-kromosoomi haplogrupilise koosseisu iseloomustamine.

### **2.2. Materjal ja metoodika**

Antud töö on osa suuremast teadusprojektist N1 fülogeograafia uurimisel. Koostöö peamiseks partneriteks on Meditsiinigeneetika Keskuse teadlased Moskvast, Biokeemia ja Geneetika Instituudi kollektiiv Ufaas ja Venemaa Teadusteakadeemia Tsütoloogia ja Geneetika Instituudi Siberi haru teadlased.

#### **2.2.1. Töös kasutatud proovid**

Käesolevas töös kasutati DNA proove Eesti Biokeskuse ja evolutsioonilise bioloogia õppetooli kollektisioonist, mis on kogutud koostöös paljude välispartneritega. Kogutud proove on varem tüpiseeritud erinevate Y-kromosoomi markerite suhtes. Antud töös kasutati 50 nganasaani, 41 neenetsi, 28 hakassi, 27 venelase, 16 tõva, 13 dolgaani, 13 tatarlase, 12 komi, 11 tšuvaši, 7 vepsa, 6 udmurdi, 6 jakuudi, 5 mari, 5 baškiiri, 5 altailase, 3 evengi, 3 sõlkupi, 3 ukrainlase, 3 mansi, 2 eenetsi, 2 valgevenelase, 2 kasahhi, 2 eskimo, 1 handi, 1 karjalase ja 1 šoori DNA proovi.

Lisaks kasutati töös 464 Y-kromosoomi, mis olid sekveneeritud Complete Genomics-i platvormil (Drmanac *et al.*, 2010, <http://www.completegenomics.com>). Saadud järjestused olid analüüsitud Complete Genomics Analysis Tools tööriistadega, kasutades referentsina GRCh37 *build*-i. Genoomsete järjestuste esialgse andmetöötamise ja konverteerimise teostasid Tartu Ülikooli Eesti geenivaramu vanemteadur Reedik Mägi, Tartu Ülikooli Eesti geenivaramu doktorant Mario Mitt, Eesti Biokeskuse vanemteadur Lauri Saag ja Eesti Biokeskuse teadur Monika Karmin.

## Y-kromosoomi analüüs

Y-kromosoomi analüüsil oli proovidel eelnevalt määratud N1b haplogruppi defineeriv marker P43 v.a nganassaanide ja neenetsite puhul, kelle Y-kromosoomid genotüpiseeriti käesolevas töös. N1b haplogruppi kuuluvatel proovidel määrati alguses mutatsiooni olemasolu positsioonis 7207924 (defineerib N1b haplogrupi Aasia spetsiifilist klaadi). Nendel proovidel, millel mutatsiooni ei olnud, määrati defineeriva mutatsiooni olemasolu positsioonis 18973691 (defineerib N1b haplogrupi Euroopa spetsiifilist klaadi).

Nganassaanide ning neenetsite Y-kromosoomi genotüpiseerimisel kasutati hierarhilise tüpiseerimise meetodit, st kõigepealt määrati proovidel marker P43, mis defineerib nendes populatsioonides kõige sagedasemat haplogruppi. Järgnevalt kontrolliti markerit TAT nendel proovidel, mis haplogruppi N1b-P43 ei kuulunud. Järgnevalt määrati TAT markeri omandatud alleeliga proovidel mutatsiooni olemasolu positsioonis 22762208, seejärel positsioonis 17216441 ning viimasena kontrolliti markerit CTS2929. Proovidel, millel oli markeri CTS2929 algne alleel, kontrolliti markerit Z1936. Proovidel, mis ei kuulunud haplogruppi N1b ega N1c, kontrolliti mutatsiooni olemasolu positsioonis 17090704. Määramata jäänud proovidel kontrolliti markerit M9. Proovidel, millel oli M9 algne alleel, määrati marker M217 ja omandatud alleeliga proovidel M242. Proovidel, millel oli M217 algne alleel, kontrolliti esmalt markerit M89 ning siis M172. Proovidel, millel oli M217 omandatud alleel, määrati kõigepealt M198 ning seejärel M458 ja M558. Proovidel, millel oli M558 algne alleel kontrolliti Z93. Mutatsioonide detekteerimiseks (markerite määramiseks) kasutati kas RFLP analüüsi (*RFLP - Restriction Fragment Length Polymorphism*) või resekveneerimist.

N1b haplogruppi kuuluvatele proovidele tehti STR analüüs, et määrata nende haplotüübid.

## 2.2.2. DNA amplifikatsioon

Y-kromosoomi amplifitseerimiseks kasutatud praimerite järjestused, polümeraasahelreaktsiooni (PCR) tingimused ja tsüklite arv on toodud lisades (vt. LISA 1 ja LISA 2). PCR-i segu komponendid ja tüvilahuste alg- ja lõppkontsentratsioonid (toodud tüvilahuse järel sulgudes) olid järgmised:

- 3 µl 5x FIREPol<sup>®</sup> MasterMix (Solis BioDyne) (koostis: FIREPol<sup>®</sup> DNA polümeraas, 5x reaktsioonipuhver B (0.4 M Tris-HCl, 0.1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% w/v Tween-20), 12.5 mM MgCl<sub>2</sub> (1x PCR lahus – 2,5 mM MgCl<sub>2</sub>), 2 mM dNTPs (1x PCR lahus – 200 µM dATP, 200 µM dCTP, 200 µM dGTP and 200 µM dTTP), sinine värv, kollane värv)
- 0,3 µl F praimer (tüvilahuse kontsentratsioon 10 pmol/µl, lõppkontsentratsioon 0,2 µM)
- 0,3 µl R praimer (tüvilahuse kontsentratsioon 10 pmol/µl, töölahuses 0,2 µM)
- 1,5-2 µl DNA
- 9,9 µl MQ
- PCR-i segu kogumaht oli 15µl.

PCR viidi läbi kaanesoojendusega termoblokis „Biometra UNO II“. Igas reaktsiooniseerias oli alati negatiivne kontroll – reaktsioonisegu ilma DNA-ta, et kontrollida kasutatud reagentide puhtust. PCR-i produkti olemasolu kontrolliti agarosgeel-elektroforeesil. 2,5%-line agarosgeel valmistati 0,5 kordse TBE puhvriga (45 mM Tris-boraat; 1 mM EDTA Na<sub>2</sub>; pH=8,3). Produkti visualiseerimiseks UV-valguses lisati geelile etiidiumbromiidi lõppkontsentratsiooniga 0,5µg/ml. Produkti kanti geelile 5 µl ja elektroforees toimus 0,5 kordses TBE puhvril horisontaalses foreesivannis konstantsel 180V pingel, elektroodide distants 32 cm. Tulemused jäädvustati UV-valguses 260 nm juures UVIPro Gold masinaga UVI Pro v12.5 tarkvara abil.



### 2.2.3. RFLP analüüs

RFLP analüüsi kasutati järgnevate markerite alleelse seisu määramiseks: P43, TAT, M9, M89, M172, M198, M207, M458, M558 ja Z93. Analüüsis kasutatud ensüümid on toodud lisas (vt. LISA 1).

Analüüsis kasutatakse restriктаase, endonukleaasse aktiivsusega ensüüme, mis lõikavad kaheaheelist DNA-d teatud spetsiifilise järjestusmotiivi esinedes. Järjestuses toimunud mutatsioonid võivad kaasa tuua äratundmissaidi kadumise või tekke. Restriksioonireaktsiooniks lisati 10 µl-le PCR-i produktile 0,3 ühikut restriктаasi (ensüümi kontsentratsiooni 5 U/µl korral 0,06 µl, ensüümi kontsentratsiooni 10 U/µl korral 0,03 µl) ja 1 µl ensüümile sobivat 10-kordset reaktsioonipuhvrit.

Restriksioon toimus 37°C juures 2-4 tundi. Restriksioonifragmentide lahutamine toimus agarosgeel-elektroforeesil (vt. peatükk 2.2.3.). Agarosgeeli kontsentratsioon oli 2,5-3,5% (sõltuvalt restriksioonifragmentide suuruselt) ja fragmentide lahutamiseks kanti geelile 10 µl restriктаasiga töödeldud PCR-i produkti.

### 2.2.4. Sekveneerimine

Sekveneerimist kasutati markerite CTS2929, Z1936, M217, M242, 7207924, 17090704, 17216441, 18973691 ja 22762208 alleelide määramiseks.

Käesolevas töös kasutati DNA sekveneerimiseks Sangeri meetodit, mis põhineb ensümaatilisel didesoksüterminatsioonil. DNA sünteesi läbiviimisel lisatakse reaktsioonisegusse DNA polümeraas, neli erinevat desoksüribonukleotiidi ning neli erinevat fluorestsentsmärgisega didesoksüribonukleotiidi, mis termineerivad ahelasse lülitudes ahela edasise sünteesi ning tekib kogum DNA fragmente, mille alguspunktiks on praimer ja lõpus on ühenukleotiidsete vahedega kõik sekveneeritava DNA ahela positsioonid. Amplifitseerimiseks kasutatatakse ainult ühte praimerit. Praimerite järjestused ja seondumistemperatuurid on toodud lisas (vt LISA 2).

Sekveneerimisreaktsioonil kasutati Applied Biosystems BigDye® Terminaator v3.1 Cycle Sequencing Kit'i.

Sekveneerimissegu kogumaht: 10 µl

- 6,14 µl MQ
- 2 µl Big Dye® Terminator v3.1 5x sekveneerimispuhver (Applied Biosystems)
- 0,7 µl Big Dye® Terminator v3.1 reagentide eelsegu (Applied Biosystems)
- 0,16 µl sekveneerimispraimer (0,16 µM)
- 1 µl PCR produkt

Sekveneerimisreaktsiooni tingimused termoblokis:

95°C	15 sek DNA denaturatsioon	} 30 tsüklit
51-58°C	10 sek praimer seondumine	
60°C	60 sek ekstensioon	

Produkti sadestamine:

- Sekveneerimisreaktsiooni produktile (10µl) lisati 2 µl NH<sub>4</sub>Ac/dekstraan segu (1 µl – 700 kDa punast dekstraani 1mg/µl; 1 µl – 7 M ammooniumatsetaat) ning 30 µl 96% (-20°C) etanooli
- Hoiti 15 minutit külmkapis temperatuuril -20°C
- Fuugiti 15 minutit kiirusel 13000 pööret minutis (p/min) (g=18 890, Hettich Zentrifugen MIKRO 22)
- Eemaldati supernatant
- Sademele lisati 200 µl 70% (-20°C) etanooli
- Fuugiti 5 minutit kiirusel 13000 p/min (g= 18 890, Hettich Zentrifugen MIKRO 22)
- Eemaldati supernatant
- Korrati pesemist 200 µl 70% (-20°C) etanooliga

- Fuugiti 5 minutit kiirusel 13000 p/min ( $g=18\ 890$ , Hettich Zentrifugen MIKRO 22)
- Eemaldati supernatant
- Lasti sademel kuivada 10 minutit  $37^{\circ}\text{C}$  kapis
- Lisati kuivanud sademele 10  $\mu\text{l}$  pealekandmislahust - 70% formamiidi (Amersham Pharmacia Biotech)
- Hoiti 30 minutit toatemperatuuril, suspendeeriti

Sekveneerimine viidi läbi Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit'i ABI PRISM® 3730xl DNA 20 analüsaatoril (Applied Biosystems).

Sekveneeritud DNA järjestusi analüüsiti programmiga ChromasPro 1.5.

### **2.2.5. STR analüüs**

Haplogruppi N1b kuuluvatel Y kromosoomidel analüüsiti 23 STR markerit: DYS19, DYS385a, DYS385b, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643, YGATAH4. Proovid pärinesid järgnevates populatsioonidest: nganassaanid (24), neenetsid (10), komid (10), tatarlased (9), dolgaanid (6), udmurdid (6), hakassid (6), vepsad (6), tšuvassid (5), sõlkupid (3), marid (2), jakuudid (2), baškiirid (2), tõvad (1), karjalased (1), evenid (1) ja evengid (1).

STR analüüsiks kasutati Promega PowerPlex® Y23 System Kit'i. Reaktsioonisegu kogumaht oli 5  $\mu\text{l}$ .

Reaktsioonisegu komponendid:

- 1  $\mu\text{l}$  PowerPlex® Y23 5x Master Mix (Promega)
- 0,5  $\mu\text{l}$  PowerPlex® Y23 10x Primer Pair Mix (Promega)
- 3  $\mu\text{l}$  MQ
- 0,5  $\mu\text{l}$  DNA-d

Reaktsioonitingimused:

96°C 2 min DNA esmane denaturatsioon

94°C 10 sek DNA denaturatsioon

61°C 1 min praimerite seondumine märklaud DNA-le

72°C 30 sek ekstensioon

60°C 20 min lõplik ekstensioon

} 30 tsüklit

Seejärel lisati sekveneerimisplaadile iga proovi kohta 9,5 µl formamiidi ja 0,35 µl CC5 Internal Lane Standardit ja siis kanti plaadile proovid. Lisaks proovidele kanti sekveneerimisplaadile 1 µl PowerPlex® Y23 Allelic Ladder Mix'i (Promega).

Alleelipikkused määrati GeneMapper® 4.0 (Applied Biosystems) tarkvara abil.

### 2.3. Praimerite disainimine

Antud töös disainiti haplogrupi N1b Euroopa alamklaadi määratleva markeri praimerid. Y-kromosoomi õige piirkonna leidmiseks vaadati järjestusi ning viidi FastA formaati *Ensembl*-i andmebaasis (<http://www.ensembl.org/index.html>). Praimerid disainiti Primer3 version 4.0.0 tööriistaga (<http://primer3.ut.ee>), kus praimeri pikkuse vahemikuks määrati 23-27 aluspaari (optimaalne 25 aluspaari). Praimeri seondumistemperatuuri vahemik oli 52-63°C (optimaalne 58°C). Monovalentsete katioonide konsentratsiooniks määrati 40 nM, divalentsete katioonide konsentratsiooniks 2,5 nM, seonduvate oligode konsentratsiooniks 300 nM ja desoksüribonukleotiidide konsentratsiooniks 0,8 nM. Disainitud praimerite seondumissaite ning PCR-i produktide hulka kontrolliti programmiga GenomeTester 1.3 (<http://bioinfo.ut.ee/genometester>). Antud programmi abil valiti välja praimerite paarid kaheksale markerile, mis olid eelnevalt välja valitud neljateistkümne haplogruppi N1b defineeriva SNP hulgast. Täiendava kontrolli, mille käigus osutusid parimateks praimerite paarid positsioonidele 16641739 (A→G) ja 18973691 (A→T), viis NCBI PrimerBLAST tööriistaga (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) läbi Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudi doktorant Anne-Mai Ilumäe. Sekveneerimistulemuste kvaliteedi põhjal otsustati kasutusse võtta praimerid positsioonile 18973691 (A→T).

## 2.4. Andmeanalüüs

Haplogrupi N fülogeneesipuu konstrueeriti 57 Y kromosoomi X-degenereerunud piirkonna järjestuse põhjal, kasutades ainult varieeruvaid nukleotiidipositsioone. Kasutati tarkvarapaketti BEAST 1.7.5 (Drummond *et al.*, 2012) (<http://beast.bio.ed.ac.uk>). Sisendfail (täiendiga .xml) tehti BEAST-i tarkvarapaketti kuuluva programmiga BEAUti. Järjestused jaotati varem teada oleva informatsiooni alusel kolme monofüleetilisse rühma. DNA järjestuse evolutsiooni mudeliks valiti GTR mudel (*GTR – general time-reversible*), milles anti ette järgmised asenduskiiruste väärtused: ac - 0.31, g - 0.98, at - 0.32, cg - 0.22 ja gt - 0.31 ja nukleotiidide sagedused: A - 0.3, C - 0.2, G - 0.2 ja T - 0.3. Võimaldamaks nukleotiidipositsioonidele erinevaid asenduskiirusi kasutati gamma jaotust. Nukleotiidipositsioonid jaotati asenduskiiruste järgi nelja rühma. Klaadide vanuse hindamisel kasutati *lognormal relaxed* kella ja mutatsioonikiirust  $2.219 \times 10^{-7}$  asendust varieeruva nukleotiidipositsiooni kohta aastas. See on saadud, teisendades mutatsioonikiiruse ( $0.82 \times 10^{-9}$  asendust nukleotiidi kohta aastas) (Poznik *et al.*, 2013), mis arvestab kõiki sekveneeritud nukleotiidipositsioone (10741501), ümber varieeruvatele nukleotiidipositsioonidele, mida 464 järjestuses oli kokku 39702. Puu arvutamiseks kasutati Eesti Biokeskuse serverit Taevas. Parima fülogeneesipuu vaatamiseks kasutati programmi FigTrees. Programmiga BEAST arvutati ka efektiivse populatsiooni suuruse dünaamika. Tulemust vaadati Programmi Tracer 1.5 abil.

Haplogrupi N1b ja selle alamklaadide geograafilise leviku kohta koostati sagedusgradientidega kaardid kasutades Surfer<sup>®</sup> 8 (Golden Software) tarkvara. Kaartide loomiseks koondati Exceli tabelisse vajalikud andmed: N1b ja alamklaadide esinemissagedused analüüsitud populatsioonides, proovide kogumispakade ligikaudsed geograafilised koordinaadid, valimite suurused ning populatsioonide nimed. Exceli fail avati Surfer<sup>®</sup> 8 programmis ja salvestati .bln failiformaadis. Sagedusgradientide kontuuride saamiseks teostati *gridding* protseduur. Saadud kontuurid viidi aluskaardile ja saadi sagedusgradientidega kaart. Kaart salvestati .srf failiformaadis ja eksporditi .jpg pildifailiks, mida töödeldi pilditöötlusprogrammiga.

139 haplogruppi N1b kuuluva Y kromosoomi fülogeneetiline võrgustik konstrueeriti 17 STR-i (DYS19, DYS385a, DYS385b, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, YS448, DYS456, DYS458, DYS635 ja YGATAH4) korduste arvu ja kaht alamklaadi defineerivate bialleelsete markerite põhjal. Mediaanvõrgustik konstrueeriti programmi Network 4.6.1.2 ja selle lisa programmi Network

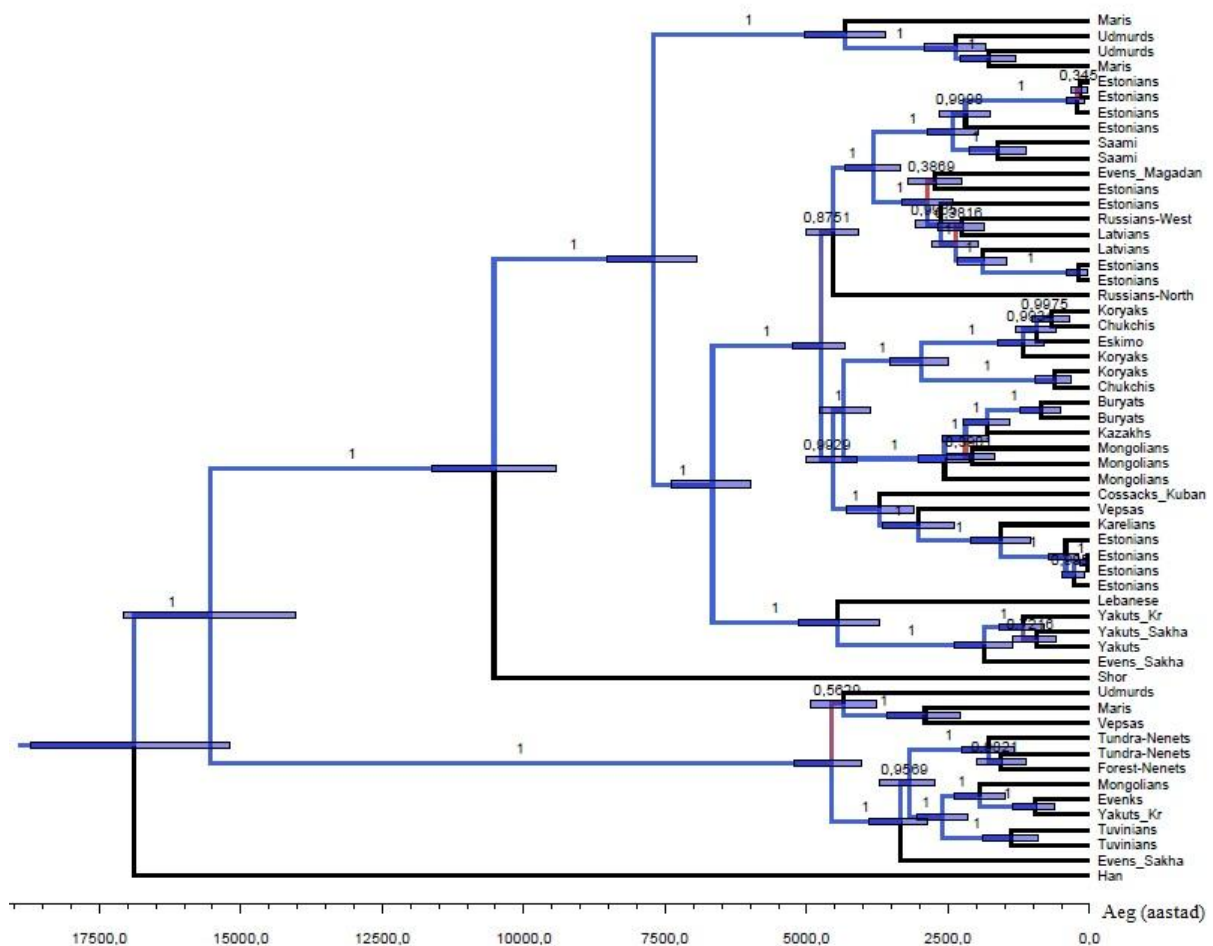
Publisher abil (A. Röhl; [www.fluxus-engineering.com](http://www.fluxus-engineering.com)), rakendades *median joining* algoritmi (Bandelt *et al.*, 1999). Mediaanvõrgu arvutamisel anti markeritele erinevad kaalud: STR-dele 1 ja SNP-dele 10.

### **3. TULEMUSED JA ARUTELU**

#### **3.1. Y-kromosoomi haplogrupi N fülogeneesipuu**

Y-kromosoomi haplogrupi N fülogeneesipuu konstrueeriti 57 sellesse haplogruppi kuuluva Y-kromosoomi järjestuse põhjal (joonis 9). Puu on statistiliselt hästi toetatud – sügavate harude toetused on maksimaalsed (aposterioorne tõenäosus 1) ning ainult üksikute tipmiste rühmade statistiline toetus on madal (aposterioorne tõenäosus  $<0,5$ ). Alamhaplogrupp N1b, mis on puul esindatud 12 järjestusega, koosneb kahest klaadist. Suurem statistiliselt hästi toetatud klaad hõlmab üheksat järjestust Siberi populatsioonidest ja on määratletud 11 SNP-ga. Väiksemat klaadi, mis koosneb vepsa ja mari järjestusest, iseloomustab 14 SNP-i. See klaad moodustab puul udmurdi järjestusega ühise rühma, mille statistiline toetus on madal, ning seetõttu võib eeldada, et udmurdi järjestus esindab pigem kolmandat klaadi.

N haplogrupi vanuse hinnanguks saadi 16900 (15200 – 18700) aastat. Haplogrupid N1b ja N1c lahknesid rohkem kui 15000 aastat tagasi ning umbes 4500 aastat tagasi algas haplogruppide N1b ja N1c oluline mitmekesisustumine.

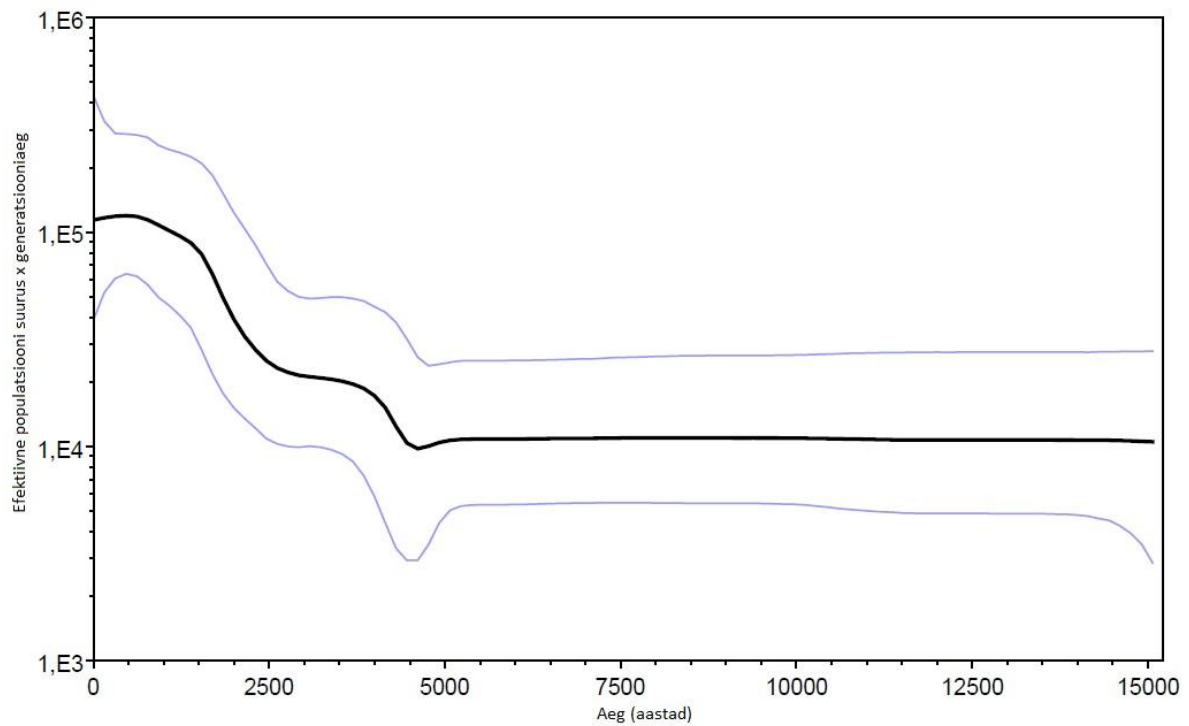


**Joonis 9.** 57 Y-kromosoomi haplogruppi N kuuluva indiviidi ~10 miljoni aluspaari pikkuste järjestuste põhjal konstrueeritud fülogeneesipuu. Puu on konstrueeritud programmiga BEAST v1.7.5. Puu harudel on toodud klaadide aposterioorsed tõenäosused. Statistiliselt hästi toetatud harud on sinised ja nõrgalt toetatud harud punased. Siniste riskülikutega on märgitud vanusehinnangute 95%-lised usaldusvahemikud. Harude tippudes on toodud populatsioonide nimed.

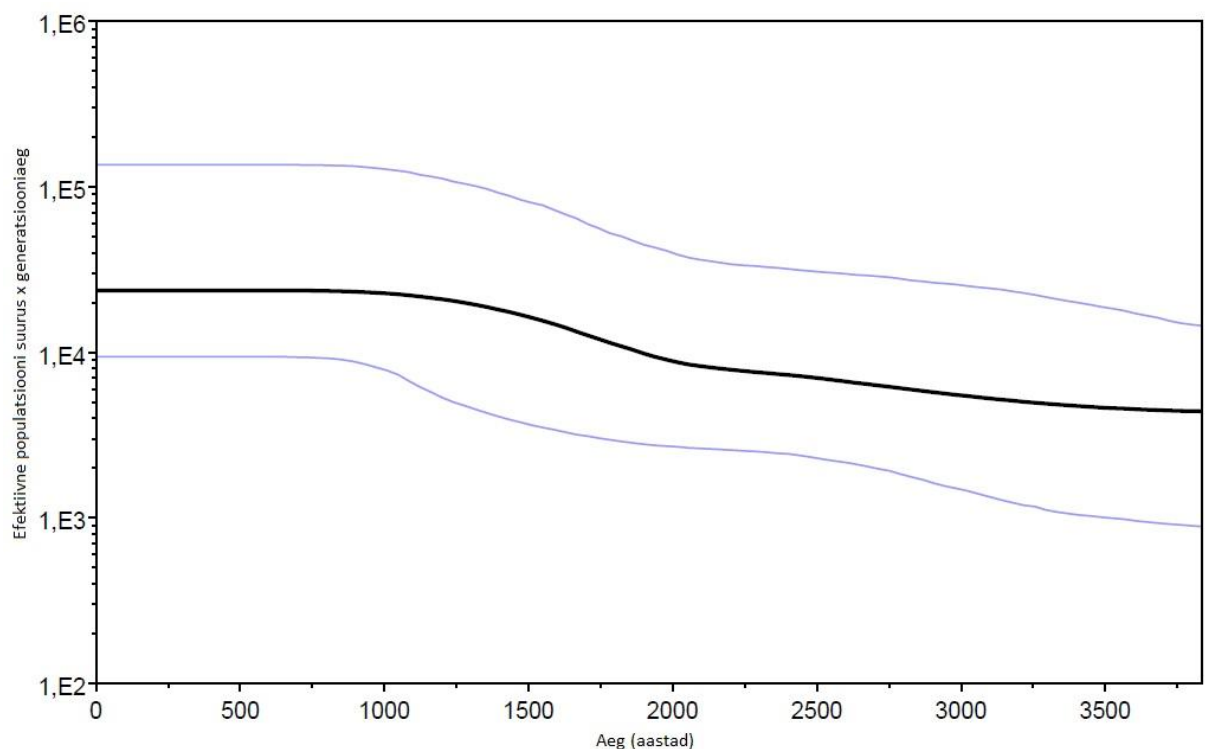
Umbes 4500 aastat tagasi hakkas populatsioon kasvama (joonis 10). Esimesele arvukuse kasvule järgnes 2500 aastat tagasi veelgi ulatuslikum efektiivse populatsiooni suuruse tõus. Üheks efektiivse populatsiooni suuruse kasvu põhjuseks võib olla haplogruppi N kandvate indiviidide migratsioon soodsamate tingimustega aladele.

Alamhaplogrupi N1b puhul on efektiivse populatsiooni suuruse muutused olnud väikesed ja toimunud aeglaselt (joonis 11). Viimase 1500 aasta jooksul on efektiivse populatsiooni suurus püsinud suhteliselt stabiilsena. See võib olla ka põhjuseks, miks N1b, võrreldes haplogrupiga N1c, on väiksema levialaga ning vähem mitmekesine.





**Joonis 10.** Efektiivse populatsiooni suuruse dünaamika. *Bayesian Skyline Plot* (BSP) on tehtud haplogrupile N 57 järjestuse põhjal, kasutades programmi BEAST v1.7.5. Mediaan on näidatud paksu musta joonega ja 95%-lised usalduspiirid siniste joontega.



**Joonis 11.** Efektiivse populatsiooni suuruse dünaamika haplogrupile N1b. *Bayesian Skyline Plot* (BSP) on tehtud 12 järjestuse põhjal, kasutades programmi BEAST v1.7.5. Mediaan on näidatud paksu musta joonega ja 95%-lised usalduspiirid siniste joontega.

### 3.2. Y-kromosoomi haplogrupi N1b fülogeograafia

Käesolevas töös määrati haplogrupi N1b ja nimetatud haplogrupi kolme alamhaplogrupi sagedused 26 Ida-Euroopa ja Siberi populatsioonis (tabel 1). N1b on kõige sagedasem neenetsitel (71%) ja Taimõri poolsaarel elavatel nganassaanidel (54%). Võrdlemisi levinud on N1b ka nganassaanidest itta jäävatel dolgaanidel (24%), Lõuna-Siberis elavatel hakassidel (16%) ja tõvadel (14%) ning Lääne-Siberis hantidel/mansidel (22%). N1b on esindatud ka Ida-Euroopas, peamiselt Volga-Uurali regiooni populatsioonidel: tatarlastel (12%), udmurtidel (11%), maridel (10%) ja tšuvašidel (10%). Keeleliselt väga lähedastel permikomidel ja sürjakomidel on N1b sagedus väga erinev vastavalt 3% ja 21%. See on ilmselt tingitud populatsioonide geograafilisest paiknemisest – sürjakomide asuala ulatub neenetsiten asulateni, aga permikomid elavad Volga-Uurali regioonis.

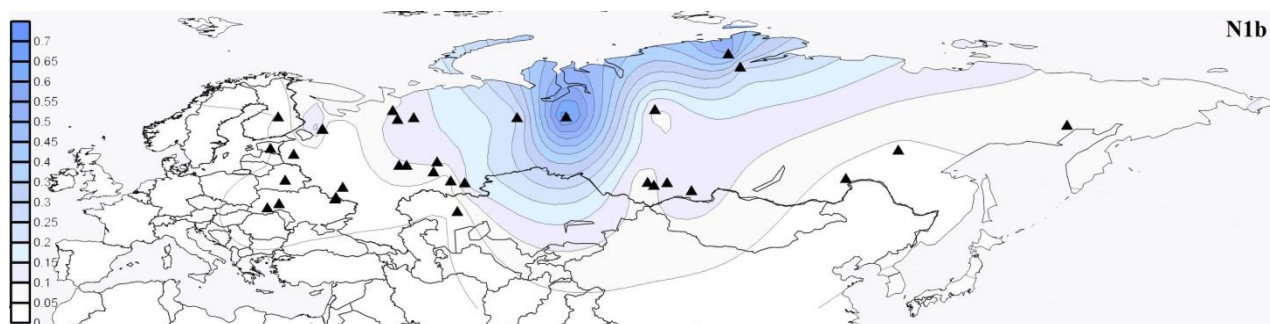
Võrreldes N1b alamhaplogruppide levikut (tabel 1), on näha, et N1b-7207924 (edaspidi Aasia-spetsiifiline klaad) on esindatud Siberi populatsioonides, olles sage nganassaanidel, neenetsitel, dolgaanidel, tõvadel ja hakassidel. Veidi vähem leidub eespool nimetatud klaadi altailastel, evenidel ja sõlkupitel. Ida-Euroopa populatsioonides (vepsad, karjalased, põhjavenelased, komid, marid, tšuvašid, tatarlased) esineb aga N1b-18973691 (edaspidi Euroopa-spetsiifiline klaad). Vähesel määral leidub Euroopa-spetsiifilist klaadi ka üksikutes Siberi populatsioonides (handid/mansid, nganassaanid, tõvad ja eskimod).

Märkimisväärne osa haplogruppi N1b kuuluvatest Y-kromosoomidest jääb parafüleetilisse rühma N1b\* ja seda eriti udmurtidel ning hantidel/mansidel. Tegelikult on olemas üks udmurdi Y-kromosoomi järjestus, mis jääb Aasia- ja Euroopa-spetsiifilistest klaadidest eraldi harusse, kuid ühe järjestuse põhjal ei saa eristada individuaalseid ja klaadi defineerivaid SNP-sid.

**Tabel 1.** Y-kromosomaalse haplogrupi N1b ja selle alamklaadide sagedused Ida-Euroopa ja Siberi populatsioonides.

<b>Populatsioon</b>	<b>Valimi suurus</b>	<b>N1b</b>	<b>N1b*</b>	<b>N1b- 7207924</b>	<b>N1b- 18973691</b>
nganassaanid	49	0,55	0,04	0,49	0,02
neenetsid	41	0,71	0,00	0,71	0
dolgaanid	54	0,24	0,02	0,22	0
venelased	594	0,05	0,01	0	0,04
sürjakomid	47	0,21	0,04	0	0,17
permikomid	60	0,03	0,02	0	0,02
udmurdid	54	0,11	0,11	0	0
marid	51	0,10	0,02	0	0,08
tõvad	95	0,14	0,00	0,12	0,02
hakassid	174	0,16	0,02	0,14	0
vepsad	40	0,20	0	0	0,20
karjalased	139	0,01	0	0	0,01
eveenid	24	0,08	0	0,08	0
evengid	57	0,04	0	0,04	0
jakuudid	157	0,04	0	0,04	0
sõlkupid	43	0,07	0	0,07	0
tšuvašid	114	0,10	0,01	0,01	0,08
tatarlased	108	0,12	0,06	0	0,06
baškiirid	568	0,01	0,01	0	0
eskimod	17	0,12	0,06	0	0,06
altailased	61	0,08	0	0,08	0
šoorid	23	0,04	0,00	0,04	0
handid/mansid	18	0,22	0,17	0	0,06
valgevenelased	267	0,01	0	0	0
ukrainlased	291	0,01	0,01	0	0
kasahhid	185	0,01	0,01	0	0

Haplogrupi N1b leviku ja sageduse iseloomustamiseks koostati antud haplogrupi (joonis 12) ja selle alamhaplogruppide sagedusgradientidega kaardid (joonis 13).



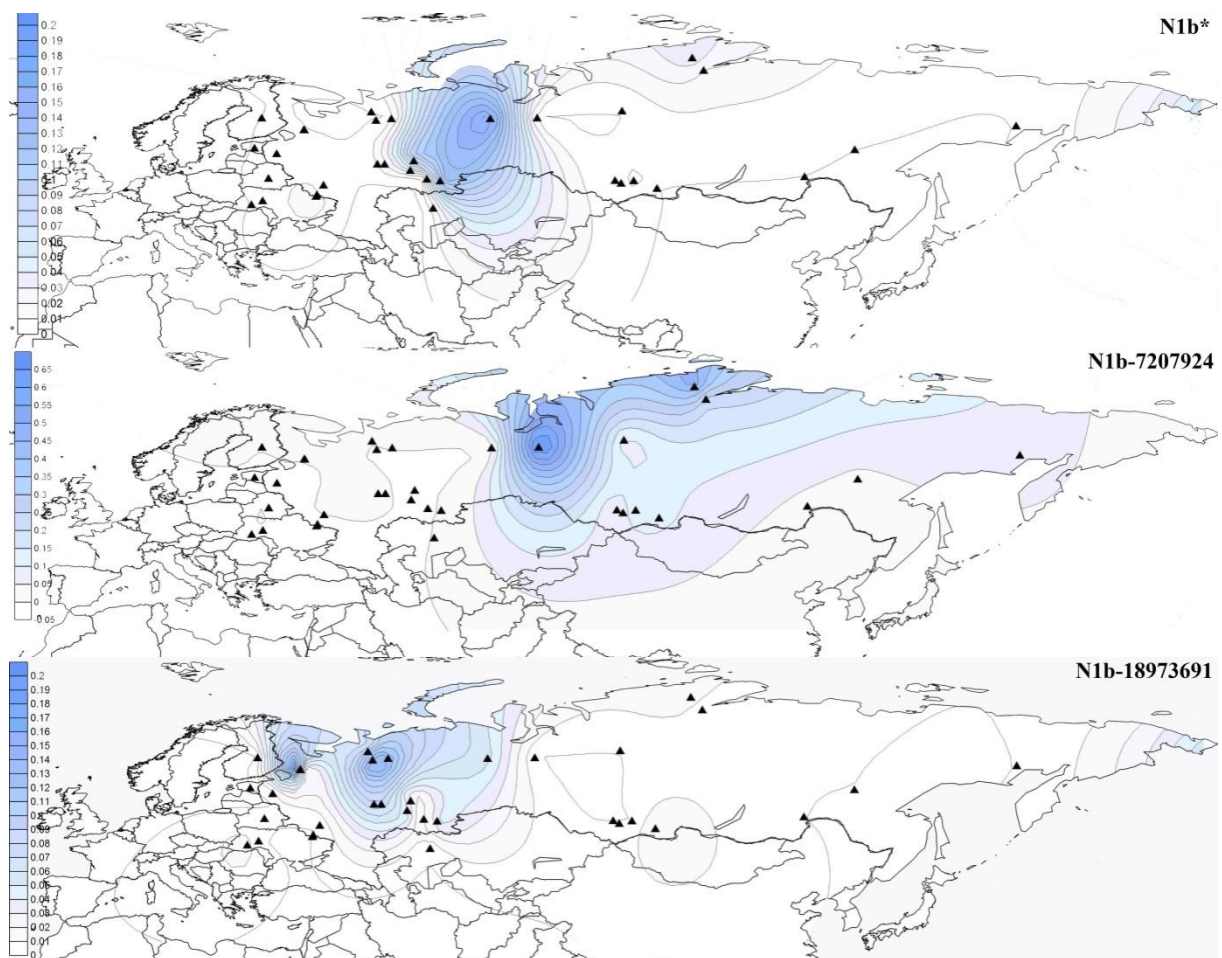
**Joonis 12.** Haplogrupp N1b-P43 geograafilise leviku ja sagedusgradientidega kaart. Kolmnurgad kaardil kujutavad analüüsitud populatsioonide ligikaudseid asukohti. Sageduse skaala on toodud vasakul.

N1b on levinud Lääne-Siberi ja Taimõri poolsaare populatsioonides, kus selle osakaal võib ulatuda isegi 70%-ni (joonis 12). Lõuna, ida ja lääne suunas haplogrupi N1b sagedus väheneb.

Kaartidelt (joonis 13) on näha, et N1b Aasia klaad on laiema levialaga kui Euroopa klaad ning sagedane just Lääne-Siberi põhjaosa populatsioonides, moodustades nendes umbes 60% haplogrupilisest mitmekesisusest. Kuigi lõuna- ja idapoolle liikudes Aasia klaadi sagedus järk-järgult väheneb, moodustab see siiski võrdlemisi märkimisväärse osa (minimaalselt 20%) sealsete populatsioonide haplogrupilisest koosseisust. Ida-Euroopa populatsioonides on Aasia haru väga marginaalne või puudub üldse.

N1b Euroopa klaad on sagedasem Põhja-Venemaa populatsioonides, kuid võrreldes Aasia haruga on Euroopa haru esinemissagedus madal – maksimaalselt 20%. Euroopa ja Aasia klaadi kokkupuuteala jääb Lääne-Siberisse.

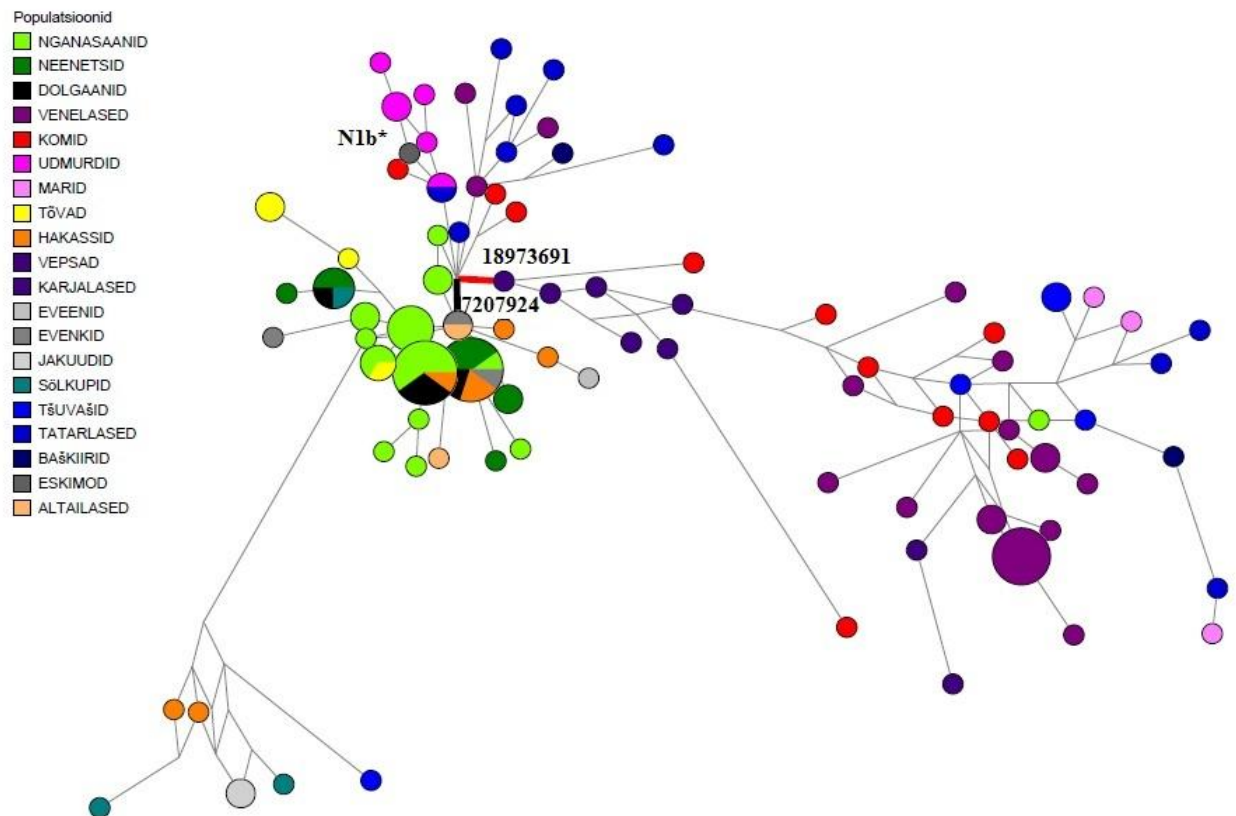
Parafüleetiline rühm N1b\* on kõige sagedasem Volga-Uurali piirkonnas udmurtidel ja Lääne-Siberis hantidel/mansidel ning sagedus väheneb lõuna, lääne ja ida suunas kiiresti.



**Joonis 13.** N1b parafüleetilise rühma N1b\*, Aasia klaadi (N1b-7207924) ja Euroopa klaadi (N1b-18973691) geograafilise leviku ning sagedusgradientidega kaart. Kolmnurgad kaardil kujutavad kasutatud populatsioonide umbkaudseid asukohti. Sageduse skaala on toodud vasakul.

### 3.3. Y-kromosoomi haplogrupi N1b haplotüüpide fülogeneetiline võrgustik

N1b alamhaplogruppe defineerivate SNP-de, antud töös 96 Y-kromosoomil analüüsitud STR-de (lisa 3) ning kirjandusest saadud 36 Y kromosoomi 17 STR-i korduste arvu (Rootsi *et al.*, 2007, Balanovsky *et al.*, 2008) põhjal koostati fülogeneetiline võrgustik (joonis 14). Võrgul eristub kolm klastrit: kaks N1b alamklaadi (Aasia- ja Euroopaspetsiifiline) ning N1b\*, millele ei ole antud hetkel leitud defineerivaid mutatsioone.



**Joonis 14.** Haplogrupi N1b haplotüüpide fülogeneetiline võrgustik. Võrgustik on konstrueeritud alamhaplogruppe defineerivate SNP-de, antud töös 96 Y-kromosoomil analüüsitud STR-de ning kirjandusest võetud 36 Y-kromosoomi 17 STR-i andmete (Rootsi *et al.*, 2007, Balanovsky *et al.*, 2008) põhjal, kasutades programmi Network 4.6.1.2 ja selle lisa programmi Network Publisher, rakendades *median joining* algoritmi. Võrgu arvutamisel anti markeritele erinevad kaalud: STR-dele 1 ja SNP-dele 10.

Aasia klaadis on kaks klastrit. Suuremasse ja juurele lähemasse rühma kuuluvad Lõuna-Siberi populatsioonid (tõvad, hakassid ja altailased), Taimõril elavad nganassaanid ja dolgaanid, Obi jõe ääres elavad neenetsid ja sõlkupid ning üksikud evenid ja evengid. Kuigi nganassaanid ja neenetsid on naabrid on neil ühiseid haplotüüpe väga vähe. Väiksemasse klastrisse langevad Ida-Siberis elavad jakuudid ning mõned sõlkupid ja hakassid.

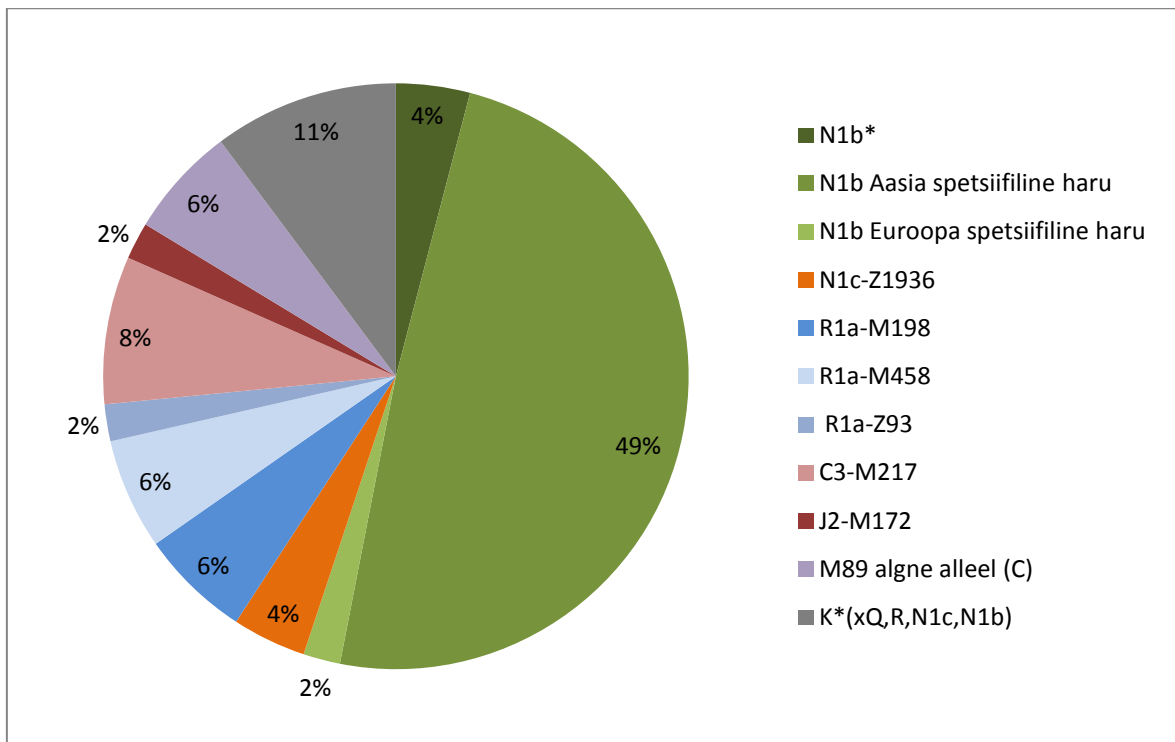
Euroopa klaadis eristub samuti kaks klastrit. Soome-ugri keeli rääkivad karjalased ja vepsad moodustavad juurele lähema klatri, millest omakorda tuleneb suurem ning arvukam rühm. Viimasesse kuuluvad Põhja-Venemaa populatsioonid (Pinega venelased ja sürjakomid) ning Volga-Uurali regiooni populatsioonidest marid, tatarlased, baškiirid ja tšuvašid.

Volga-Uurali regiooni keskne parafüleetiline rühm N1b\* jaguneb samuti kaheks klastriks, millest ühte kuuluvad peamiselt udmurdid, teise aga tšuvaššid, tatarlased ja venelased.

### **3.4. Nganassaanide ja neenetsite Y-kromosoomi haplogrupiline koosseis**

Kahe Lääne-Siberis elava uurali keeli kõneleva populatsiooni – nganassaanide ja neenetsite – Y-kromosoomi haplogrupilise koosseisu iseloomustamiseks genotüpiseeriti 49 nganassaanide ja 41 neenetsi Y kromosoomid.

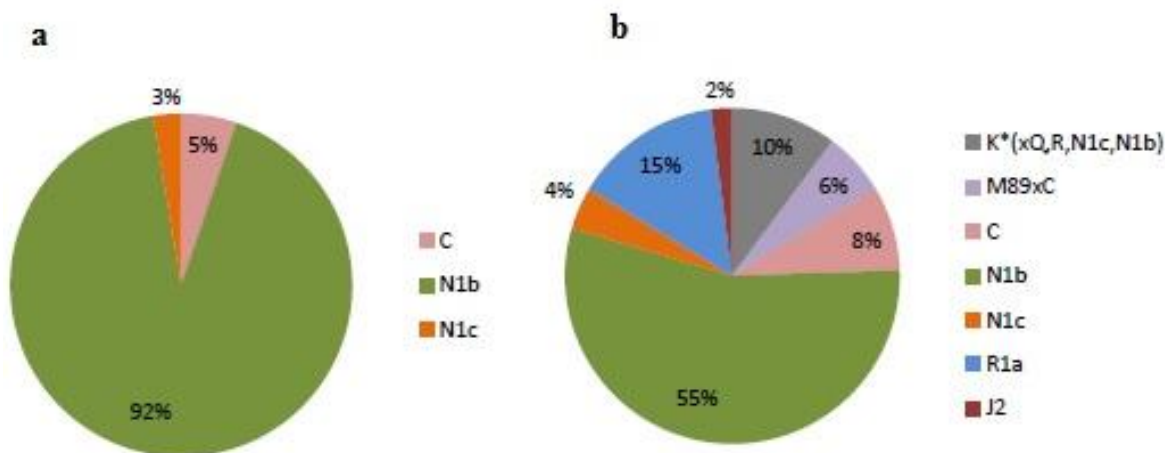
Analüüsitud nganassaanide Y-kromosoomidest kuulub 49% haplogrupi N1b (N-P43) Aasia-spetsiifilisse klaadi, 2% Euroopa-spetsiifilisse klaadi, ja 4% jääb N1b parafüleetilisse rühma (N1b\*) (joonis 15). Haplogrupi N1c klaadis N1c-Z1936 on 4% proovidest. Umbes 8% Y kromosoomidest kuulub haplogruppi C3 (C3-M217). Küllaltki suur osa jääb R1a kolme alamhaplogruppi – 6% R1a-M198, 6% R1a-M458 ja 2% R1a-Z93, milledest R1a-M458 on levinud enamasti Euroopa populatsioonides ja R1a-Z93 Aasias. Umbes 2% Y kromosoomidest kuulub haplogruppi J2-M172. DNA halva kvaliteedi tõttu jäi lõpuni genotüpiseerimata kaheksa proovi, millest kolmel on markeri M89 algne alleel (C) ja viis proovi kuuluvad haplogruppi K\*(xQ,R,N1c,N1b). Analüüsi tulemustest on hästi näha, et nganassaanid jäävad Ida- ja Lääne-Euraasia geenitiigi kokkupuute ja segunemise alasse. Suur osa (~67%) nganassaanide Y-kromosoomidest kuulub haplogruppidesse N1b, N1c ja C3, mis on eeldatavasti tekkinud Ida-Aasias ning sealt läände levinud, samas kui Lääne-Euraasiale iseloomulik haplogrupp R1a moodustab 15% analüüsitud Y kromosoomidest.



**Joonis 15.** Y-kromosomaalsete haplogrupid osakaalud nganassanide populatsioonis

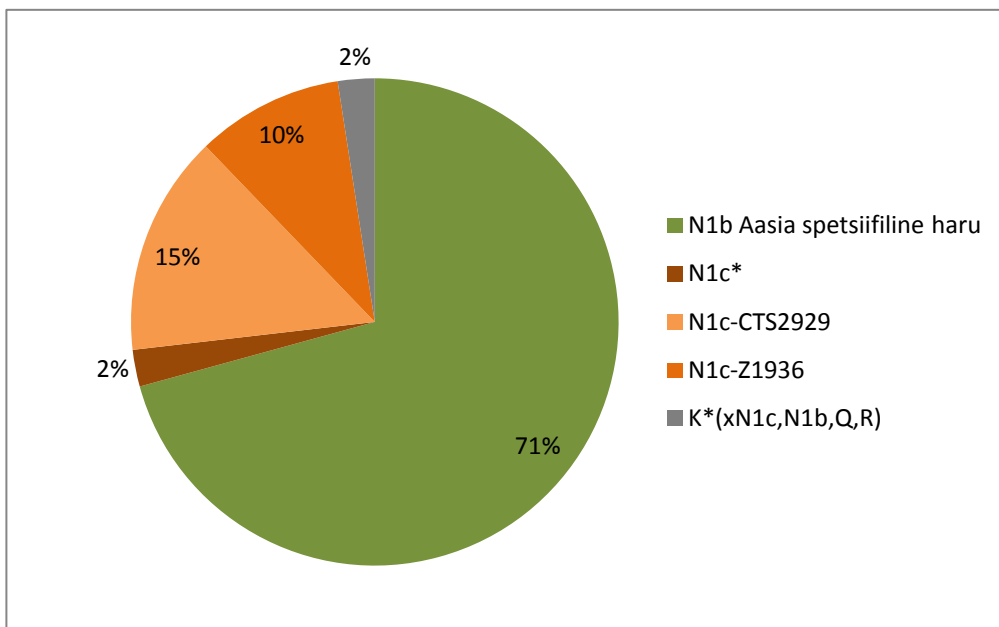
Võrdlemaks käesoleva töö tulemusi kirjanduses avaldatud andmetega (Karafet *et al.*, 2002) koostati kaks diagrammi, kusjuures paremaks võrdlemiseks viidi antud töö andmed kirjanduse andmetega samale genotüpiseerimise tasemele (joonis 16). Käesoleva töö nganassaanide valim on palju mitmekesisem – esindatud on ka haplogrupid (R1a, J2). Mõlema valimi puhul on levinuimaks haplogrupiks N1b, kuid sagedused erinevad märkimisväärselt – 92% (Karafet *et al.*, 2002) ja käesolevas töös 55%. Kirjanduse andmete järgi on sageduselt järgmine haplogrupp C (5%), kuid antud töös R1a (15%). Haplogrupi C sagedus on käesolevas töös 8%. Haplogrupp N1c on mõlemas valimis esindatud sarnase sagedusega (3-4%). Erinevuste põhjuseks võivad olla väikesed valimid (38 indiviidi kirjandusest ja 49 indiviidi käesolevas töös). Lisaks võib erinevusi põhjustada asjaolu, et proovid ei ole kogutud samadest paikkondadest.





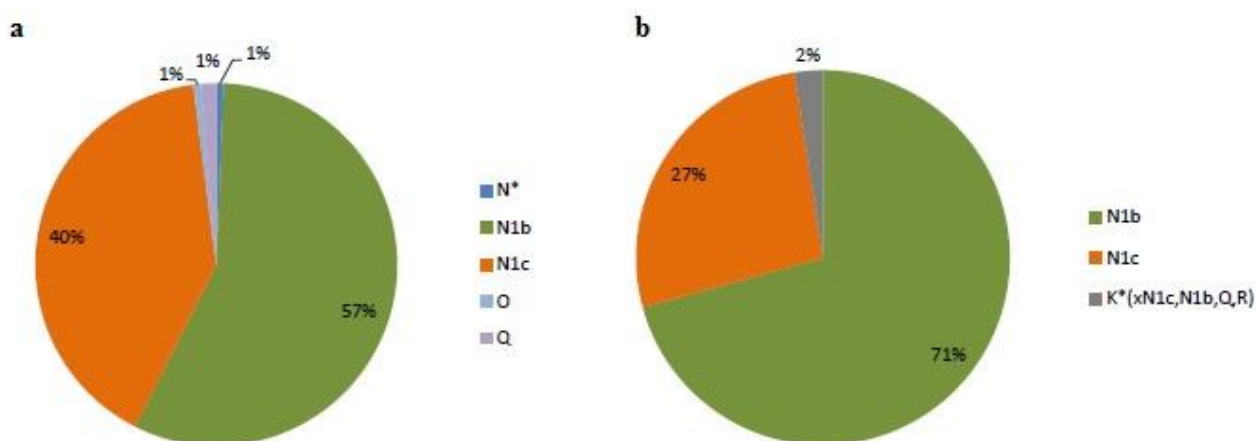
**Joonis 16.** Y-kromosomaalsete haplogruppide sageduste võrdlus nganassaanidel. Joonisel a on kujutatud sagedused, mis põhinevad kirjandusest saadud andmetel (Karafet *et al.*, 2002), ja joonisel b on sagedused, mis põhinevad käesoleva töö andmetel.

Neenetsite Y kromosoomi geenitiik on võrreldes nganassaanidega palju ühetaolisem (joonis 17). Sarnaselt naaberpopulatsioonile kuulub enamus (~71%) neenetside haplogruppi N1b Aasia-spetsiifilisse alamhaplogruppi. Antud valimi ülejäänud proovid kuuluvad N1c kolme alamhaplogruppi: N1c-CTS2929 (15%), N1c-Z1936 (10%) ja N1c\* (2%). Üks proov jäi haplogruppi K\*(xN1c,N1b,Q,R).



**Joonis 17.** Neenetsite Y-kromosoomide haplogrupiline jaotuvus.

Ka neenetsitel võrreldi käesoleva analüüsi tulemust kirjanduses avaldatud andmetega (Karafet *et al.*, 2002) (joonis 18). Kahe andmestiku võrdlemiseks viidi antud töö genotüpiseerimise tulemused kirjanduses avaldatutega samale genotüpiseerimise tasemele. Kuigi valimite suurused erinesid märkimisväärselt (avaldatud töös 148 ja käesolevas töös 41 indiviidi), on neenetsitel kahe valimi haplogrupiline koostis sarnasem kui nganassaanidel. Siiski, peamiste haplogruppide (N1b ja N1c) sagedused on valimiti erinevad. Avaldatud andmetes on sagedused vastavalt 58% ja 40% ning antud töös vastavalt 71% ja 27%. Erinevus on tõenäoliselt põhjustatud asjaolust, et metsa neenetsitel on N1c haplogrupp palju sagedasem kui tundra neenetsitel ja kirjanduses avaldatud valimis moodustavad suure osa (60%) just metsa neenetsid, aga käesoleva töö valimist on enamuse tundra neenetsid.



**Joonis 18.** Y-kromosomaalsete haplogruppide sageduste võrdlus neenetsitel. Joonisel a on kujutatud sagedused, mis põhinevad kirjandusest saadud andmetel (Karafet *et al.*, 2002), ja joonisel b on sagedused, mis põhinevad käesoleva töö andmetel.

Tuleb ära mainida, et tulemusi võib mõjutada võrdlemisi väike valim, mis on tihti tingitud populatsioonide väiksusest ning ka proovide raskest kättesaadavusest.

## KOKKUVÕTE

Käesolevas töös käsitleti Euraasia põhjaosas levinud Y-kromosomaalse haplogrupi N osa uurali keeli kõnelevate populatsioonide geenitiigi kujunemisel. Haplogrupi paremaks iseloomustamiseks konstrueeriti 57 Y-kromosoomi järjestuse põhjal N haplogrupi fülogeneesipuu, millelt hinnati ka haplogruppide vanuseid. N haplogrupi kujunemine jääb 17000 aasta taha ning alamhaplogrupid N1c ja N1b lahkesid umbes 15000 aastat tagasi. Veidi enam kui 4000 aastat tagasi hakkas efektiivse populatsiooni suurus kasvama ning haplogruppides N1c ja N1b tekkis mitmeid uusi klaade. Teine palju ulatuslikum mitmekesisustumine ja populatsiooni kasv algas umbes 2500 aastat tagasi. Üheks efektiivse populatsiooni suuruse kasvu põhjuseks võis olla haplogruppi N kandvate indiviidide migratsioon soodsamate tingimustega aladele.

Haplogrupi N1b fülogeograafia iseloomustamiseks määrati nimetatud haplogrupi ja selle alamhaplogruppide sagedused 26 Ida-Euroopa ja Siberi populatsioonis. Tulemuste visualiseerimiseks koostati sagedusgradientide kaardid. N1b on enim esindatud Lääne-Siberi ja Taimõri poolsaare populatsioonides, kus sagedus ulatub 70%-ni. Aasia-spetsiifiline klaad on esindatud Siberi populatsioonides, olles sage nganassaanidel, neenetsitel, dolgaanidel, tõvadel ja hakassidel. Ida-Euroopa populatsioonides on Aasia haru marginaalne või puudub üldse. Euroopa-spetsiifiline haru esineb Ida-Euroopa populatsioonides, saavutades suurima sageduse vepsadel ning sürjakomidel. Vähesel määral leidub Euroopa-spetsiifilist klaadi ka üksikutes Siberi populatsioonides. Kahe klaadi kokkupuuteala jääb Lääne-Siberisse. Parafüleetiline rühm N1b\* on kõige sagedasem Volga-Uurali piirkonnas udmurtidel ja Lääne-Siberis hantidel/mansidel ning sagedus väheneb lõuna, lääne ja ida suunas kiiresti.

Haplogrupi N1b kolme alamhaplogrupi sisemise varieeruvuse iseloomustamiseks koostati alamhaplogruppe defineerivate SNP-de ja 132 Y kromosoomi 17 STR-i korduste arvu põhjal fülogeneetiline võrgustik. Aasia-spetsiifilises klaadis eristub kaks klastrit, millest juurele lähemasse ja arvukamasse kuuluvad Lõuna-Siberi populatsioonid, Taimõril elavad nganassaanid ja dolgaanid ning Obi jõe äärsed neenetsid ja sõlkupid. Euroopa-spetsiifilises klaadis moodustavad Soome-ugri keeli rääkivad karjalased ning vepsad juurele lähemal paikneva klatri, millest tuleneb suurem klaster, kuhu kuuluvad põhjavenelased ning esindajaid mitmest Volga-Uurali regiooni populatsioonist. Parafüleetilises rühmas N1b\* on enim esindatud Volga-Uurali regiooni populatsioonid. Ühe klatri moodustavad peamiselt udmurdid ja teise tšuvaššid, tatarlased ja venelased.

Kahe Lääne-Siberis elava uurali keeli kõneleva populatsiooni – nganassaanide ja neenetsite – Y-kromosoomi haplogrupilist koosseisu iseloomustati 49 genotüpiseeritud nganassaanide ja 41 neenetsi Y kromosoomi alusel. Nganassaanid on geneetiliselt mitmekesisemad, kui nende naabruses elavad samasse keelkonda kuuluvad neenetsid. Nganassaanidel tuvastati järgmised Y-kromosomaalsed haplogrupid: N1b, N1c, R1a, C3 ja J2. Neenetsitel esineb eelnevalt nimetatud haplogruppidest vaid N1b ja N1c. Nii nganassaanidel kui ka neenetsitel kuulub enamus proovidest Y-kromosomaalse haplogrupi N1b Aasia-spetsiifilisse harusse, vastavalt 55% ja 71%. Haplogrupp N1c on neenetsitel sagedasem kui nganassaanidel, kuid neenetsitel puudub haplogrupp R1a, mis nganassaanidel moodustab märkimisväärse osa haplogrupilisest koosseisust.

Käesolev töö moodustab ühe osa suuremast teadusprojektist N1 fülogeograafia uurimisel.

# **The role of Y-chromosomal haplogroup N in the formation of genepool of Uralic-speaking populations**

Helen Post

## **SUMMARY**

Male specific Y chromosome which is inherited from father to son has been the basis of phylogenetic and –geographic studies for years. A big part of the Y chromosome called male specific region does not recombine with the X chromosome and therefore it passes intact from one generation to the other. In past studies of the Y chromosome mainly single nucleotide polymorphisms and short tandem repeats have been used. Lately because of the fast development of sequencing methods the sequences of Y chromosome X-degenerate region have been used to estimate mutation rate, to refine the topology of the phylogenetic tree and to find new haplogroup defining markers.

The main aim of this study was to describe the inner structure of Y chromosomal haplogroup N and its role in the formation of the genepool of Uralic-speaking populations. The composition of Y chromosomal haplogroups was more thoroughly characterised in two Uralic-speaking populations – the Nganasans and the Nenets.

Using 57 Y chromosome sequences and the BEAST v1.7.5. software an N haplogroup phylogenetic tree was constructed. Based on this tree the time estimates were found for two haplogroups – N and N1b – ~17 000 and ~ 15 000 years, respectively. Additionally, the dynamics of the effective population size for both haplogroups was assessed. It seems that the first extensive variation of the N haplogroup took place about 4500 years ago and the second even larger expansion took place about 2000 years later.

To describe the phylogeography of Y chromosomal haplogroup N1b the subclades – Asian specific, European specific – of the said haplogroup were determined in samples from 26 different populations inhabiting East Europe and Siberia. Asian specific clade was most frequent in West Siberia (up to 71%) and South-Siberia (up to 14%). European specific clade is all in all less frequent than the Asian specific clade and more common in East Europe (~20%). Notable amount of Y chromosomes mostly from Volga-Uralic populations belong to the paraphyletic group. In addition, a median-joining network was constructed using the SNP and 17 STR data of 132 N1b haplogroup samples. The clustering on the network verifies that the most frequent N1b subclade in Western and Southern Siberian populations is Asian

specific and the most frequent N1b subclade in Eastern European is European specific but also paraphyletic group.

The Y chromosomal haplogroup composition in the Nganasans is more diversified than in their geographic neighbours the Nenets. The Nganasan samples belonged to N1b, N1c, R1a, C3 and J2 haplogroups. The Nenets samples belonged mainly to N1b ja N1c. Most samples from both populations subjected to haplogroup N1b and its Asian specific subclade – the frequency in the Nenets was 71% and in the Nganasans 55%.

# TÄNUAVALDUSED

Tahan tänada oma juhendajaid Maere Reidlat ja Siiri Rootsit suurepärase juhendamise, igakülgse abi ja heade nõuannete eest. Samuti tahan tänada Eesti Biokeskuse vanemteadurit Lauri Saagi, kes aitas N haplogrupi fülogeneesipuu konstrueerimisel ja eriti programmi BEAST kasutamisel ning Tartu Ülikooli Molekulaar- ja Rakubioloogia Instituudi doktoranti Anne-Mai Ilumäed, kes aitas praimerite disainil. Lisaks tänan Evolutsioonilise bioloogia õppetooli inimesi toetuse, abivalmiduse ning meeldiva tööõhkkonna eest.

Käesolev magistritöö on saanud toetust Genoomika tippkeskuse uurimistoetuselt 3.2.0101.08-0011.

## KASUTATUD KIRJANDUS

Bachtrog D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 14: 113-124.

Balanovsky O., Rootsi S., Pshenichnov A., Kivisild T., Churnosov M., Evseeva I., Pocheshkhova E., Boldyreva M., Yankovsky N., Balanovska E., Villems R. (2008). Two sources of the Russian patrilineal heritage in their Eurasian context. *The American Journal of Human Genetics* 82: 236-250.

Bandelt H. J., Forster P., Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37-48.

Burton M. L., Moore C. C., Whiting J. W. M., Romney K. (1996). Regions Based on Social Structure. *Current Anthropology* 37: 87-123.

Cole J. E. 2011. Udmurts, p. 371-372. In J. E. Cole (ed.), *Ethnic Groups of Europe: An Encyclopedia*, ABC-CLIO, Santa Barbara.

Cruciani F., Trombetta B., Massaia A., Destro-Bisol G., Sellitto D., Scozzari R. (2011). A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *The American Journal of Human Genetics* 88: 814-818.

Cui Y., Li H., Ning C., Zhang Y., Chen L., Zhao X., Hagelberg E., Zhou H. (2013). Y Chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. *BMC Evolutionary Biology* 13: 216.

de Knijff P. (2000). Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *The American Journal of Human Genetics* 67: 1055-1061.

Derenko M., Malyarchuk B., Denisova G. A., Wozniak M., Dambueva I., Dorzhu C., Luzina F., Miścicka-Sliwka D., Zakharov I. (2006). Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Human Genetics* 118: 591-604.

Derenko M., Malyarchuk B., Denisova G., Wozniak M., Grzybowski T., Dambueva I., Zakharov I. (2007). Y-chromosome haplogroup N dispersals from south Siberia to Europe. *Journal of Human Genetics* 52: 763-770.



Drmanac R., Sparks A. B., Callow M. J., Halpern A. L., Burns N. L., Kermani B. G., Carnevali P., Nazarenko I., Nilsen G. B., Yeung G., Dahl F., Fernandez A., Staker B., Pant K. P., Baccash J., Borcharding A. P., Brownley A., Cedeno R., Chen L., Chernikoff D., Cheung A., Chirita R., Curson B., Ebert J. C., Hacker C. R., Hartlage R., Hauser B., Huang S., Jiang Y., Karpinchyk V., Koenig M., Kong C., Landers T., Le C., Liu J., McBride C. E., Morenzoni M., Morey R. E., Mutch K., Perazich H., Perry K., Peters B. A., Peterson J., Pethiyagoda C. L., Pothuraju K., Richter C., Rosenbaum A. M., Roy S., Shafto J., Sharanhovich U., Shannon K. W., Sheppy C. G., Sun M., Thakuria J. V., Tran A., Vu D., Zaranek A. W., Wu X., Drmanac S., Oliphant A. R., Banyai W. C., Martin B., Ballinger D. G., Church G. M., Reid C. A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 5961: 78-81.

Drummond A. J., Suchard M. A., Xie D., Rambaut A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969-1973.

Duggan A. T., Whitten M., Wiebe V., Crawford M., Butthof A., Spitsyn V., Makarov S., Novgorodov I., Osakovsky V., Pakendorf B. (2013). Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One* 12: e83570.

Dulik M. C., Osipova L. P., Schurr T. G. (2011). Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One* 6: e17548.

Dulik M. C., Zhadanov S. I., Osipova L. P., Askapuli A., Gau L., Gokcumen O., Rubinstein S., Schurr T. G. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *The American Journal of Human Genetics* 90: 229-246.

Elhaik E., Tatarinova T. V., Klyosov A. A., Graur D. (2014). The ‘extremely ancient’ chromosome that isn’t: a forensic bioinformatic investigation of Albert Perry’s X-degenerate portion of the Y chromosome. *European Journal of Human Genetics*: doi: 10.1038/ejhg.2013.303.

Fedorova S. A., Reidla M., Metspalu E., Metspalu M., Rootsi S., Tambets K., Trofimova N., Zhadanov S. I., Hooshiar Kashani B., Olivieri A., Voevoda M. I., Osipova L. P., Platonov

F. A., Tomskey M. I., Khusnutdinova E. K., Torroni A., VILLEMS R. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evolutionary Biology* 13: 127.

FrancaLacci P., Morelli L., Angius A., Berutti R., Reinier F., Atzeni R., Pili R., Busonero F., Maschio A., Zara I., Sanna D., Useli A., Urru M. F., Marcelli M., Cusano R., Oppo M., Zoledziewska M., Pitzalis M., Deidda F., Porcu E., Poddie F., Kang H. M., Lyons R., Tarrier B., Gresham J. B., Li B., Tofanelli S., Alonso S., Dei M., Lai S., Mulas A., Whalen M. B., Uzzau S., Jones C., Schlessinger D., Abecasis G. R., Sanna S., Sidore C., Cucca F. (2013). Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science* 341: 565-569.

Golovnev A. V. 1999. The Nia (Nganasan), p. 166-169. In R. B. Lee and R. Daly (ed.), *The Cambridge Encyclopedia of Hunters and Gatherers*, 1<sup>st</sup> ed., Cambridge University Press, Cambridge.

Golovnev A. V., Osherenko G. 1999. Introduction: „The People“ or „the Others“, p. 1-15. In A. V. Golovnev (ed.), *Siberian Survival. The Nenets and Their Story*, 1<sup>st</sup> ed., Cornell University Press, New York.

Graves J. A. (2006). Sex chromosome specialization and degeneration in mammals. *Cell* 124: 901-914.

Heyer E., Puymirat J., Dieltjes P., Bakker E., de Knijff P. (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Molecular Genetics* 6: 799-803.

Honkola T., Vesakoski O., Korhonen K., Lehtinen J., Syrjänen K., Wahlberg N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* 26: 1244-1253.

Hughes J. F., Rozen S. (2012). Genomics and genetics of human and primate Y chromosomes. *Annual Review of Genomics and Human Genetics* 13: 83-108.

Jobling M. A., Tyler-Smith C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* 8: 598-612.

- Karafet T. M., Osipova L. P., Gubina M. A., Posukh O. L., Zegura S. L., Hammer M. F. (2002). High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Human Biology* 74: 761-789.
- Karafet T. M., Mendez F. L., Meilerman M. B., Underhill P. A., Zegura S. L., Hammer M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830-838.
- Karlsson A. O., Wallerström T., Götherström A., Holmlund G. (2006). Y-chromosome diversity in Sweden - a long-time perspective. *European Journal of Human Genetics* 14: 963-970.
- Kharkov V. N., Stepanov V. A., Feshchenko S. P., Borinskaia S. A., Iankovskii N. K., Puzyrev V. P. (2005). Frequencies of Y Chromosome binary haplogroups in belarussians. *Journal of Russian Genetics* 41: 928-931.
- Lachance J., Vernot B., Elbers C. C., Ferwerda B., Froment A., Bodo J. M., Lema G., Fu W., Nyambo T. B., Rebbeck T. R., Zhang K., Akey J. M., Tishkoff S. A. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African Hunter-Gatherers. *Cell* 150: 457-469.
- Lappalainen T., Koivumäki S., Salmela E., Huoponen K., Sistonen P., Savontaus M. L., Lahermo P. (2006). Regional differences among the Finns: a Y-chromosomal perspective. *Gene* 376: 207-215.
- Lappalainen T., Laitinen V., Salmela E., Andersen P., Huoponen K., Savontaus M. L., Lahermo P. (2008). Migration waves to the Baltic Sea region. *Annals of Human Genetics* 72: 337-348.
- McDougall I., Brown F. H., Fleagle J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 7027: 733-736.
- Mendez F. L., Krahn T., Schrack B., Krahn A. M., Veeramah K. R., Woerner A. E., Fomine F. L., Bradman N., Thomas M. G., Karafet T. M., Hammer M. F. (2013). An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *The American Journal of Human Genetics* 92: 454-459.

Mirabal S., Regueiro M., Cadenas A. M., Cavalli-Sforza L. L., Underhill P. A., Verbenko D. A., Limborska S. A., Herrera R. J. (2009). *European Journal of Human Genetics* 17: 1260-1273.

Olson J. S. 1994. Komi, p. 390-392. In J. S. Olson, L. Pappas, and N. C. J. Pappas (ed.), *An Ethnohistorical Dictionary of the Russian and Soviet Empires*, 1st ed. Greenwood Press, Westport.

Oota H., Settheetham-Ishida W., Tiwawech D., Ishida T., Stoneking M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nature Genetics* 29: 20-21.

Passarino G., Cavalleri G. L., Lin A. A., Cavalli-Sforza L. L., Børresen-Dale A. L., Underhill P. A. (2002). Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *European Journal of Human Genetics* 10: 521-529.

Pimenoff V. N., Comas D., Palo J. U., Vershubsky G., Kozlov A., Sajantila A. (2008). Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics* 16: 1254-1264.

Poznik G. D., Henn B. M., Yee M. C., Sliwerska E., Euskirchen G. M., Lin A. A., Snyder M., Quintana-Murci L., Kidd J. M., Underhill P. A., Bustamante C. D. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341: 562-565.

Rootsi S., Magri C., Kivisild T., Benuzzi G., Help H., Bermisheva M., Kutuev I., Barac L., Pericic M., Balanovsky O., Pshenichnov A., Dion D., Grobei M., Zhivotovsky L. A., Battaglia V., Achilli A., Al-Zahery N., Parik J., King R., Cinnioğlu C., Khusnutdinova E., Rudan P., Balanovska E., Scheffrahn W., Simonescu M., Brehm A., Goncalves R., Rosa A., Moisan J. P., Chaventre A., Ferak V., Füredi S., Oefner P. J., Shen P., Beckman L., Mikerezi I., Terzić R., Primorac D., Cambon-Thomsen A., Krumina A., Torroni A., Underhill P. A., Santachiara-Benerecetti A. S., Villems R., Semino O. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *The American Journal of Human Genetics* 75: 128-137.

Rootsi S., Zhivotovsky L. A., Baldovic M., Kayser M., Kutuev I. A., Khusainova R., Bermisheva M. A., Gubina M., Fedorova S. A., Ilumäe A. M., Khusnutdinova E.

K., Voevoda M. I., Osipova L. P., Stoneking M., Lin A. A., Ferak V., Parik J., Kivisild T., Underhill P. A., Villems R. (2007). A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15: 204-211.

Rootsi S., Behar D. M., Järve M., Lin A. A., Myres N. M., Passarelli B., Poznik G. D., Tzur S., Sahakyan H., Pathak A. K., Rosset S., Metspalu M., Grugni V., Semino O., Metspalu E., Bustamante C. D., Skorecki K., Villems R., Kivisild T., Underhill P. A. (2013). Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nature Communications* 4: 2928.

Sanchez J. J., Børstinga C., Hernandezb A., Mengel-Jørgensen J., Morlinga N. (2003). Y chromosome SNP haplogroups in Danes, Greenlanders and Somalis. *Progress in Forensic Genetics* 10 1261: 347-349.

Semino O., Passarino G., Oefner P. J., Lin A. A., Arbuzova S., Beckman L. E., De Benedictis G., Francalacci P., Kouvatsi A., Limborska S., Marcikiae M., Mika A., Mika B., Primorac D., Santachiara-Benerecetti A. S., Cavalli-Sforza L. L., Underhill P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 5494: 1155-1159.

Shi H., Qi X., Zhong H., Peng Y., Zhang X., Ma R. Z., Su B. (2013). Genetic evidence of an East Asian origin and Paleolithic northward migration of Y-chromosome Hhaplogroup N. *PLoS One* 8: e66102.

Skaletsky H., Kuroda-Kawaguchi T., Minx P. J., Cordum H. S., Hillier L., Brown L. G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S. F., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfling T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S. P., Waterston R. H., Wilson R. K., Rozen S., Page D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825-837.

Syrjänen K., Honkola T., Korhonen K., Lehtinen J., Vesakoski O., Wahlberg N. (2013). Shedding more light on language classification using basic vocabularies and phylogenetic methods. A case study of Uralic. *Diachronica* 30: 323-352.

Zerjal T., Dashnyam B., Pandya A., Kayser M., Roewer L., Santos F. R., Schiefenhövel W., Fretwell N., Jobling M. A., Harihara S., Shimizu K., Semjidsmaa D., Sajantila A., Salo P., Crawford M. H., Ginter E. K., Evgrafov O. V., Tyler-Smith C. (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *The American Journal of Human Genetics* 60: 1174-1183.

Zhivotovsky L. A., Underhill P. A., Cinnioglu C., Kayser M., Morar B., Kivisild T., Scozzari R., Cruciani F., Destro-Bisol G., Spedini G., Chambers G. K., Herrera R. J., Yong K. K., Gresham D., Tournev I., Feldman M. W., Kalaydjieva L. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *The American Journal of Human Genetics* 74: 50-61.

Zhivotovsky L. A., Underhill P. A., Feldman M. W. (2006). Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Molecular Biology and Evolution* 23: 2268-2270.

Tambets K., Rootsi S., Kivisild T., Help H., Serk P., Loogväli E. L., Tolk H. V., Reidla M., Metspalu E., Pliss L., Balanovsky O., Pshenichnov A., Balanovska E., Gubina M., Zhadanov S., Osipova L., Damba L., Voevoda M., Kutuev I., Bermisheva M., Khusnutdinova E., Gusar V., Grechanina E., Parik J., Pennarun E., Richard C., Chaventre A., Moisan J. P., Barac L., Pericić M., Rudan P., Terzić R., Mikerezi I., Krumina A., Baumanis V., Koziel S., Rickards O., De Stefano G. F., Anagnou N., Pappa K. I., Michalodimitrakis E., Ferák V., Füredi S., Komel R., Beckman L., Villems R. (2004). The western and eastern roots of the Saami—the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *The American Journal of Human Genetics* 74: 661-682.

Underhill P. A., Poznik G. D., Rootsi S., Järve M., Lin A. A., Wang J., Passarelli B., Kanbar J., Myres N. M., King R. J., Di Cristofaro J., Sahakyan H., Behar D. M., Kushniarevich A., Sarac J., Saric T., Rudan P., Pathak A. K., Chaubey G., Grugni V., Semino O., Yepiskoposyan L., Bahmanimehr A., Farjadian S., Balanovsky O., Khusnutdinova E. K., Herrera R. J., Chiaroni J., Bustamante C. D., Quake S. R., Kivisild T., Villems R. (2014). The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics Online publication*: doi: 10.1038/ejhg.2014.50.

Underhill P. A., Kivisild T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Reviews of Genetics* 41: 539-564.

van Oven M., Van Geystelen A., Kayser M., Decorte R., Larmuseau M. H. (2013). Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Human Mutation* 35: 187-191.

Wei W., Ayub Q., Chen Y., McCarthy S., Hou Y., Carbone I., Xue Y., Tyler-Smith C. (2012). A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Research* 23: 388-395.

Wells R. S., Yuldasheva N., Ruzibakiev R., Underhill P. A., Evseeva I., Blue-Smith J., Jin L., Su B., Pitchappan R., Shanmugalakshmi S., Balakrishnan K., Read M., Pearson N. M., Zerjal T., Webster M. T., Zholoshvili I., Jamarjashvili E., Gambarov S., Nikbin B., Dostiev A., Aknazarov O., Zalloua P., Tsoy I., Kitaev M., Mirrakhimov M., Chariev A., Bodmer W. F. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 98: 10244-10249.

Xue Y., Wang Q., Long Q., Ng B. L., Swerdlow H., Burton J., Skuce C., Taylor R., Abdellah Z., Zhao Y., Asan, MacArthur D. G., Quail M. A., Carter N. P., Yang H., Tyler-Smith C. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology* 19: 1453-1457.

YCC. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12: 339-348.

## **KASUTATUD VEEBIAADRESSID**

**<http://www.ethnologue.com>** - Ethnologue, Languages of the World

**<http://www.eki.ee/knab/linguae.htm>** - Maailma keeled, kirjad ja rahvad, versioon 1.5 (16.09.2013)

**<http://www.britannica.com>** – Encyclopaedia of Britannica, online

**<http://www.ensembl.org/index.html>** - Ensembl andmebaas

**<http://primer3.ut.ee>** - Primer3web version 4.0.0, Pick primers from a DNA sequence

**<http://bioinfo.ut.ee/genometester>**- GenomeTester 1.3

**<http://www.fluxus-engineering.com>**- Fluxus Technology Ltd, programm Network 4.6.1.2

**<http://beast.bio.ed.ac.uk>** – tarkvarapakett Beast 1.7.5

**<http://www.ncbi.nlm.nih.gov/tools/primer-blast>** - NCBI, PrimerBlast

**<http://harpending.humanevo.utah.edu/popstr>** - programm POPSTR

**<http://www.isogg.org/tree>** - Y-DNA Haplogroup Tree 2014, International Society of Genetic Genealogy, versioon: 9.63 kuupäev: 14 May 2014

**<http://www.completegenomics.com>** – Complete Genomics

**<http://www.eki.ee/books/redbook/>** - The Red Book of the Peoples of the Russian Empire (Vene Impeeriumi rahvaste punane raamat), Margus Kolga, Igor Tõnurist, Lembit Vaba, Jüri Viikberg, 1993.



## LISAD

**LISA 1.** Kasutatud SNP-markerite praimerid, PCR-i ja RFLP analüüsi tingimused.

Markeri nimetus (mutatsioon)	Praimeri järjestus 5'-3'		PCR-i tingimused	PCR-i tsüklite arv	Produkti pikkus	Restriktaas	Fragmentide pikkused	Viide
P43 (C→T)	F	TTTGGAGGGACATTA TTCTC	1. 94°C 3 min, 2. 94°C 30 sek,	35	519 ap	Hin1II	251 ap, 268 ap	YCC, 2002
	R	GAAGCAATACTCTGA AAAGT	3. 53°C 30 sek, 4. 72°C 30 sek, 5. 72°C 3 min					
TAT (T→C)	F	GACTCTGAGTGTAGA CTTGTGA	1. 94°C 3 min, 2. 94°C 15 sek,	34	112 ap	Hin1II	85 ap, 27 ap	Zerjal <i>et al.</i> , 1997
	R	GAAGGTGCCGTAAA AGTGTGAA	3. 60°C 15 sek, 4. 72°C 20 sek, 5. 72°C 3 min					
M9 (C→G)	F	GCAGCATATAAAACT TTCAGGA	1. 94°C 3 min, 2. 94°C 15 sek,	34	127 ap	Hinf1	68 ap, 59 ap	Underhill <i>et al.</i> , 1997
	R	GTAAGACATTGAACG TTTGAACA	3. 50°C 20 sek, 4. 72°C 20 sek, 5. 72°C 3 min					
M89 (C→T)	F	CATGAAAGTGGGGC CCACAG	1. 94°C 3 min, 2. 94°C 30 sek,	35	103 ap	N1a III	79 ap, 24 ap	Underhill <i>et al.</i> , 2001
	R	GCAACTCAGGCAAAA GTGAGACAT	3. 58°C 30 sek, 4. 72°C 45 sek, 5. 72°C 3 min					

(jätkub)

# LISA 1. Järg

Markeri nimetus (mutatsioon)	Praimeri järjestus 5'-3'		PCR-i tingimused	PCR-i tsükli arv	Produkti pikkus	Restriktaas	Fragmentide pikkused	Viide
M458 (A→G)	F	AGAAGAGATTTCTA GCCAGAGT	1. 94°C 3 min, 2. 94°C 30 sek,	35	219 ap	Hin1 II	198 ap, 21 ap	Underhill <i>et al.</i> , 2009
	R	CCAGAAAAAATAA TTCCTGGA	3. 53°C 30 sek, 4. 72°C 30 sek, 5. 72°C 3 min					
M558 (T→C)	F	TGTTGGCTGGCCTCT CTC	1. 94°C 3 min, 2. 94°C 25 sek,	37	211 ap	Eco471	T 172 ap, 11ap	Underhill <i>et al.</i> , 2014
	R	GAACAAGGCAGTTG TAGGATAG	3. 55°C 30 sek, 4. 72°C 35 sek, 5. 72°C 3 min				C 110 ap, 97 ap, 75 ap	
Z93 (G→A)	F	AACAAAGCATCATC AAAGGC	1. 94°C 3 min, 2. 94°C 25 sek,	36	338 ap	AluI	G 182 ap, 111 ap, 45 ap	Magoon G., „1000 Genomes Project“
	R	CATGATTCGTTATGA CCTGC	3. 57°C 25 sek, 4. 72°C 30 sek, 5. 72°C 3 min				A 128 ap, 111 ap, 54 ap, 45 ap	

(jätkub)

## LISA 1. Järg

Markeri nimetus (mutatsioon)		Praimeri järjestus 5'-3'	PCR-i tingimused	PCR-i tsükli arv	Produkti pikkus	Restri ktaas	Fragmentide pikkused	Viide
M207 (A-G)	F	AGGAAAAATCAGAAGT ATCCCTG	1. 94°C 3 min, 2. 94°C 20 sek,	37	423 ap	DraI	346 ap, 77ap	Underhill et al., 2001
	R	CAAAATTCACCAAGAA TCCTTG	3. 53°C 20 sek, 4. 72°C 35 sek, 5. 72°C 3 min					
M172 (T→G)	F	TTGAAGTTACTTTTATA ATCTAATGCTT	1. 94°C 3 min, 2. 94°C 30 sek,	35	173 ap	HinFI	145 ap, 28 ap	Underhill et al., 2001
	R	ATAATTTATTACTTTAC AGTCACAGTGG	3. 53°C 30 sek, 4. 72°C 30 sek, 5. 72°C 3 min					
M198 (C→T)	F	TGAGGTGGAATGTATC AGTATACC	1. 94°C 3 min, 2. 94°C 30 sek,	36	444 ap	SacI	398 ap, 46 ap	Underhill <i>et al.</i> , 2001
	R	TGATTTCAAGGATTTG TTAGTCTT	3. 53°C 35 sek, 4. 72°C 40 sek, 5. 72°C 3 min					

**LISA 2.** Sekveneerimisel kasutatud SNP-markerite praimerid ja PCR-i tingimused.

Markeri nimetus (mutatsioon)	Praimeri järjestus 5'-3'		PCR-i tingimused	PCR-i tsüklite arv	Produkti pikkus	Viide
CTS2929 <sup>5</sup> (T→C)	F	TGGACATATACACCCAC ACT	1. 94°C 3 min, 2. 94°C 25 sek 3. 58°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	228	avaldamata
	R	GATTGGGGAAAAGTTGG TCA				
Z1936 <sup>5</sup> (C→T)	F	CTAAACTCGTCCCTCAGT CA	1. 94°C 3 min, 2. 94°C 25 sek 3. 58°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	232	avaldamata
	R	GCTTAACTCTGCCTGACT TC				
7207924 <sup>4</sup> (T→A)	F	CAAGAGCAAGGTCTAGT GTA	1. 94°C 3 min, 2. 94°C 25 sek 3. 57°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	214	avaldamata
	R	ACTCTCTACCCTCTGCAA AA				
M217 <sup>2</sup> (A→C)	F	GCTTATTTTGTAGTCTCTC TTCCAT	1. 94°C 3 min, 2. 94°C 30 sek 3. 54°C 35 sek, 4. 72°C 40 sek, 72°C 3 min	38	461	Underhill <i>et al.</i> , 2001
	R	ACCTGTTGAATGTTACAT TTCTTT				
M242 <sup>1</sup> (T→C)	F	AACTCTTGATAAACCGT GCTG	1. 94°C 3 min, 2. 94°C 30 sek 3. 53°C 30 sek, 4. 72°C 35 sek, 72°C 3 min	36	366	Cinnioglu <i>et al.</i> , 2004
	R	TCCAATCTCAATTCATGC CTC				
17090704 <sup>3</sup> (G→A)	F	TGTTTCTAGTTGCCCTGA TG	1. 94°C 3 min, 2. 94°C 25 sek 3. 56°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	196	avaldamata
	R	AGATGACAGACGGACCT TAA				
17216441 <sup>6</sup> (G→C)	F	AGAGTGTGTTTCTTTACT GCT	1. 94°C 3 min, 2. 94°C 25 sek 3. 60°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	197	avaldamata
	R	GACACAGGAAGCATGAC AA				

(jätkub)

## LISA 2. Järg

Markeri nimetus (mutatsioon)	Praimeri järjestus 5'-3'		PCR-i tingimused	PCR-i tsüklite arv	Produkti pikkus	Viide
18973691 <sup>6</sup> (A→T)	F	ATTATGTCTGCGTTGTCTTATTG AA	1. 94°C 3 min, 2. 94°C 25 sek 3. 60°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	391	avaldamata
	R	GAGACTCAAGCCTGTAATTTTGA A				
22762208 <sup>4</sup> (T→A)	F	AAGAAACCTAAGAAAGCCTGC	1. 94°C 3 min, 2. 94°C 25 sek 3. 57°C 25 sek, 4. 72°C 25 sek, 72°C 3 min	38	250	avaldamata
	R	ATGTAAGAACGTGCTATCTG				

Markeri nimetuse või positsiooni taga on ülanumbritega märgitud sekveneerimisreaktsiooni seondumistemperatuurid vastavalt: 1- 51°C, 2- 52°C, 3- 54°C, 4- 55°C, 5- 56°C ja 6- 58°C.

**LISA 3.** Antud töös määratud proovide N1b-P43 alamklaadid ja 23 STR markerit.

ID	popula- sioon	P43 (C-T)	7207924 (T-A)	1897369 1 (A-T)	DY S19	DYS 385a	DYS 385b	DYS 389 I	DYS3 89 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS 448	DYS 456	DYS 458	DYS 481	DYS 533	DYS 549	DYS 570	DYS 576	DYS 635	DYS 643	YGA TAH4
nga8 6A	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	15	21	12	13	17	19	24	10	12
nga8 1A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	16	24	9	12
nga7 8A	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	19	24	10	12
nga7 1A	nganas aanid	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	15	17	21	13	13	18	17	25	10	12
nga4 4A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga4 0A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	19	24	10	12
nga2 33A	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga1 A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	18	15	16	21	12	13	17	18	24	10	11
nga1 89A	nganas aanid	T	A		14	12	12	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	18	18	24	10	12
nga1 80A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	19	24	10	12
nga1 67A	nganas aanid	T	T	A	14	12	13	12	16	23	10	14	13	14	10	10	19	15	17	21	12	12	17	17	24	10	12
nga1 38A	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga1 2A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	18	15	16	22	12	12	17	19	24	10	12
nga1 18A	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga8 8B	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga7 9B	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	11	19	15	16	21	12	13	18	18	24	10	12
nga6 5B	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	11	19	15	16	21	12	13	17	19	24	10	12
nga6 3B	nganas aanid	T	A		14	13	13	13	16	24	10	14	13	14	10	10	18	15	16	21	12	13	17	18	24	10	12
nga4 1B	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
nga1 99B	nganas aanid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	18	18	24	10	12
nga1 90B	nganas aanid	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	15	21	12	13	17	19	24	10	12

jätkub

### LISA 3. Järg

ID	popula tsioon	P43 (C-T)	720792 4 (T-A)	1897369 1 (A-T)	DY S19	DYS 385a	DYS 385b	DYS 389 I	DYS3 89 II	DY S39 0	DY S39 1	DY S39 2	DY S39 3	DY S43 7	DY S43 8	DY S43 9	DY S44 8	DY S45 6	DY S45 8	DY S48 1	DY S53 3	DY S54 9	DY S57 0	DY S57 6	DY S63 5	DY S64 3	YGA TAH4
nen1 15	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	14	18	17	25	9	12
nen1 27	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	20	12	13	17	16	24	9	12
nen1 33	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	16	24	9	12
nen1 12	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	18	17	25	9	12
nen1 45	neenet sid	T	A		14	12	13	13	16	23	11	14	13	14	10	9	19	15	16	22	12	13	17	17	24	11	12
nen1 60	neenet sid	T	A		14	12	13	13	16	23	11	14	13	14	10	9	19	15	16	22	12	13	17	17	24	11	12
nen2 10	neenet sid	T	A		14	13	13	13	15	23	10	14	13	14	10	11	19	15	17	21	12	13	17	17	24	9	12
nen2 24	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	17	24	10	12
nen2 28	neenet sid	T	A		14	12	13	13	16	23	11	14	13	14	10	9	19	15	16	22	12	13	17	17	25	11	12
nen1 63	neenet sid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	17	24	10	12
dolg DA75	dolgaa nid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	18	18	24	10	12
dolg DD48	dolgaa nid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	16	24	9	12
dolg DD53	dolgaa nid	T	A		14	12	13	13	16	23	11	14	13	14	10	9	19	15	16	22	12	13	17	17	24	11	12
dolg DD60	dolgaa nid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	20	12	13	18	17	24	10	12
dolg DD85	dolgaa nid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	17	16	24	9	12
dolg DD94	dolgaa nid	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	17	18	24	10	12
kpvi1 0	komid	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	15	16	22	13	12	17	17	26	11	12
kpvi8 9	komid	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	15	16	22	12	12	17	18	25	11	12
kpvi9 1	komid	T	T	T	13	12	12	13	17	23	10	12	13	14	10	10	18	15	16	22	13	12	17	18	25	11	12
kpvs 40	komid	T	T	T	13	12	13	13	18	23	10	12	13	14	10	10	19	15	15	21	13	13	17	18	24	10	12

jätkub

### LISA 3. Järg

ID	popula tsioon	P43 (C-T)	7207924 (T-A)	1897369 1 (A-T)	DY S19	DYS 385a	DYS 385b	DYS 389 I	DYS3 89 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS 448	DYS 456	DYS 458	DYS 481	DYS 533	DYS 549	DYS 570	DYS 576	DYS 635	DYS 643	YGA TAH4
kpvs 41	komid	T	T	A	14	12	13	14	16	23	10	14	13	14	10	10	19	14	17	22	12	13	19	18	24	10	12
kpvs 52	komid	T	T	T	13	12	12	13	17	23	10	12	13	14	10	10	18	15	15	21	13	13	17	18	24	10	12
kpvs 69	komid	T	T	A	14	12	13	14	16	24	10	14	13	14	10	10	19	15	17	21	12	13	18	18	24	10	12
kpvs 75	komid	T	T	T	13	12	12	13	17	23	10	12	13	14	10	10	19	15	16	21	13	13	17	17	23	10	12
kpvs 76	komid	T	T	T	14	12	13	15	16	23	10	14	13	14	10	11	19	15	18	21	12	13	19	18	24	10	12
koi7 0	komid	T	T	A	14	12	14	13	16	23	10	15	13	14	11	10	19	15	17	21	12	13	19	17	24	10	12
koi1 14	komid	T	T	T	14	11	11	13	16	24	10	15	13	14	11	10	20	15	15	23	12	13	18	18	23	10	13
udm 72	udmur did	T	T	A	14	12	13	13	16	23	10	15	13	14	11	10	19	15	17	21	12	13	19	17	22	10	12
udm 78	udmur did	T	T	A	14	12	13	13	16	23	10	15	13	14	11	10	19	15	17	21	12	14	19	17	23	10	12
udm 95	udmur did	T	T	A	14	12	13	13	16	24	10	15	13	14	11	10	19	15	17	21	13	14	19	18	23	10	12
udm 104	udmur did	T	T	A	14	12	13	13	16	24	10	15	13	14	11	10	19	15	16	21	13	14	19	17	23	10	12
udm 108	udmur did	T	T	A	14	12	13	13	16	23	10	15	13	14	11	10	19	15	17	21	12	14	18	17	24	10	12
udm 110	udmur did	T	T	A	14	12	13	13	16	24	10	15	13	14	11	10	19	15	17	21	12	14	19	18	23	10	12
mrj2 44	marid	T	T	T	13	12	12	13	18	23	10	12	13	14	10	11	18	15	18	21	12	13	17	18	23	10	12
mrj5 45	marid	T	T	T	13	12	12	13	16	23	10	12	13	14	10	10	18	16	20	22	13	13	16	17	24	10	12
tuva 29	tõvad	T	A		15	12	13	13	16	23	11	14	13	14	10	10	19	15	16	21	12	14	17	18	24	10	12
khak 99	hakassi	T	A		14	12	13	14	16	23	10	14	14	14	10	10	19	15	17	21	12	13	17	18	24	10	12
khak 144	hakassi	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	18	17	24	11	12
khak 173	hakassi	T	A		13	11	12	13	18	23	10	11	12	14	10	10	18	15	16	21	12	13	17	18	24	10	12
khak 179	hakassi	T	A		13	11	12	13	18	23	10	12	12	14	10	10	18	15	16	21	12	12	15	18	25	10	12

jätkub



### LISA 3. Järg

ID	popula tsioon	P43 (C-T)	7207924 (T-A)	1897369 1 (A-T)	DY S19	DYS 385a	DYS 385b	DYS 389 I	DYS3 89 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS 448	DYS 456	DYS 458	DYS 481	DYS 533	DYS 549	DYS 570	DYS 576	DYS 635	DYS 643	YGA TAH4
khak 192	hakassi	T	A		14	12	13	13	16	23	10	14	13	14	10	10	19	15	18	21	12	13	17	17	24	10	12
khak 195	hakassi	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	18	18	24	10	12
veps 3	vepsad	T	T	T	13	12	13	13	18	23	10	11	12	14	10	10	18	16	16	21	12	13	17	17	23	10	12
veps 33	vepsad	T	T	T	14	12	13	13	16	23	11	14	13	14	10	10	19	15	15	21	12	13	17	19	23	10	12
veps 36	vepsad	T	T	T	14	12	13	13	16	23	10	15	13	14	10	10	19	15	15	21	12	13	17	17	23	10	12
veps 45	vepsad	T	T	T	14	12	13	13	16	23	10	14	13	14	10	10	19	15	15	23	12	12	17	17	24	10	12
veps 57	vepsad	T	T	T	14	12	13	13	16	23	10	14	13	14	10	10	19	15	16	22	12	13	17	17	24	10	12
veps 63	vepsad	T	T	T	14	12	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	18	17	24	10	12
kare l86	karjalas ed	T	T	T	14	13	13	13	16	23	10	14	13	14	10	10	19	15	16	21	12	13	18	18	23	12	12
en1 5	eveeni d	T	A		14	12	13	14	16	23	10	14	14	14	10	10	19	15	17	21	12	13	17	18	24	10	11
ek31	evenki d	T	A		14	13	13	13	16	23	10	14	13	14	10	10	19	15	17	21	12	13	18	17	24	10	12
ny3 9	jakuudi d	T	A		13	11	12	13	18	23	10	11	12	14	11	10	18	15	16	21	12	13	16	18	23	10	12
Cy 17	jakuudi	T	A		13	11	12	13	18	23	10	11	12	14	11	10	18	15	16	21	12	13	16	18	23	10	12
selT K37	sõlkupi d	T	A		13	12	13	13	18	23	10	11	12	14	11	10	18	15	16	21	12	13	16	18	23	10	12
selP 32	sõlkupi d	T	A		13	12	13	13	18	23	10	11	12	14	10	10	18	15	16	21	12	13	17	18	25	10	12
selK P4	sõlkupi d	T	A		14	12	13	13	16	23	11	14	13	14	10	9	19	15	16	22	12	13	17	17	24	11	12
cju3 0	tšuvaši d	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	14	18	21	12	13	19	18	23	10	12
cju3 3	tšuvaši d	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	15	18	22	12	14	16	17	25	10	12
cju4 5	tšuvaši d	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	14	18	21	12	13	17	18	23	11	12
cju6 3	tšuvaši d	T	T	T	13	12	12	13	18	23	10	12	13	14	10	10	18	15	16	21	13	13	17	18	24	10	12

jätkub

### LISA 3. Järg

ID	populatsioon	P43 (C-T)	7207924 (T-A)	1897369 1 (A-T)	DY S19	DYS 385a	DYS 385b	DYS 389 I	DYS3 89 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS 448	DYS 456	DYS 458	DYS 481	DYS 533	DYS 549	DYS 570	DYS 576	DYS 635	DYS 643	YGA TAH4
cju91	tšuvašid	T	A		13	12	12	13	18	23	10	12	13	14	10	10	18	14	18	21	12	13	18	18	23	10	12
ttrk252	tatarlased	T	T	A	14	12	13	13	16	23	10	14	13	14	10	10	18	15	16	21	12	12	15	17	25	11	12
bsxa40	baškiirid	T	T	T	13	12	12	13	17	23	10	12	13	14	10	10	18	15	19	22	13	13	16	17	25	10	12
TAT T10	tatarlased	T	T	T	13	12	13	13	18	22	10	12	13	14	10	10	18	15	18	22	12	14	17	17	24	10	12
TAT T20	tatarlased	T	T	A	14	11	12	13	16	23	10	12	12	14	10	10	18	15	15	21	13	12	17	17	24	10	12
TAT T26	tatarlased	T	T	A	14	12	13	14	16	23	10	13	12	14	10	10	18	15	16	20	12	12	17	16	24	10	12
TAT T30	tatarlased	T	T	T	14	12	13	13	17	23	10	12	13	14	10	10	18	15	18	23	12	12	17	17	24	10	12
TAT T33	tatarlased	T	T	A	15	12	13	13	16	23	10	14	13	14	10	10	18	15	17	21	12	12	15	17	25	10	12
TAT T39	tatarlased	T	T	A	14	12	13	13	16	23	10	13	13	14	10	10	18	15	16	21	13	12	17	16	25	10	12
TAT T51	tatarlased	T	T	A	14	12	13	13	16	23	10	14	13	14	10	10	19	15	17	21	11	14	18	18	24	10	13
tatk70	tatarlased	T	T	A	14	12	13	13	16	23	10	15	13	14	11	10	19	15	17	21	12	14	20	18	24	10	12
burz15	baškiirid	T	T	A	14	12	12	13	16	23	10	14	13	14	10	10	18	15	16	21	12	12	17	18	23	10	12

## **Lisades kasutatud artiklid**

Cinnioğlu C., King R., Kivisild T., Kalfoğlu E., Atasoy S., Cavalleri G. L., Lillie A. S., Roseman C. C., Lin A. A., Prince K., Oefner P. J., Shen P., Semino O., Cavalli-Sforza L. L., Underhill P. A. (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Annals of Human Genetics* 65: 43-62.

Zerjal T., Dashnyam B., Pandya A., Kayser M., Roewer L., Santos F. R., Schiefenhövel W., Fretwell N., Jobling M. A., Harihara S., Shimizu K., Semjidsmaa D., Sajantila A., Salo P., Crawford M. H., Ginter E. K., Evgrafov O. V., Tyler-Smith C. (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *The American Journal of Human Genetics* 60: 1174-1183.

Underhill P. A., Jin L., Lin A. A., Mehdi S. Q., Jenkins T., Vollrath D., Davis R. W., Cavalli-Sforza L. L., Oefner P. J. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* 10: 996-1005.

Underhill P. A., Passarino G., Lin A. A., Shen P., Mirazón Lahr M., Foley R. A., Oefner P. J., Cavalli-Sforza L. L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics* 65: 43-62.

Underhill P. A., Myres N. M., Rootsi S., Metspalu M., Zhivotovsky L. A., King R. J., Lin A. A., Chow C. E., Semino O., Battaglia V., Kutuev I., Järve M., Chaubey G., Ayub Q., Mohyuddin A., Mehdi S. Q., Sengupta S., Rogaev E. I., Khusnutdinova E. K., Pshenichnov A., Balanovsky O., Balanovska E., Jeran N., Augustin D. H., Baldovic M., Herrera R. J., Thangaraj K., Singh V., Singh L., Majumder P., Rudan P., Primorac D., Vilems R., Kivisild T. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *European Journal of Human Genetics* 18: 479-484.

Underhill P. A., Poznik G. D., Rootsi S., Järve M., Lin A. A., Wang J., Passarelli B., Kanbar J., Myres N. M., King R. J., Di Cristofaro J., Sahakyan H., Behar D. M., Kushniarevich A., Sarac J., Saric T., Rudan P., Pathak A. K., Chaubey G., Grugni V., Semino O., Yepiskoposyan L., Bahmanimehr A., Farjadian S., Balanovsky O., Khusnutdinova E. K., Herrera R. J., Chiaroni J., Bustamante C. D., Quake S. R., Kivisild T., Vilems R. (2014). The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics Online publication*: doi: 10.1038/ejhg.2014.50.

YCC. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12: 339-348.

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina Helen Post

(sünnikuupäev: 30.04.1990)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

Y-kromosomaalse haplogrupi N osa uurali keeli kõnelevate populatsioonide geenitiigi kujunemisel,

mille juhendajad on Maere Reidla ja Siiri Rootsi,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates **26.05.2017** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 26.05.2014 (*kuupäev*)